EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT

Volume II

1942

Published by

SCIENCE RESEARCH ASSOCIATES

1700 PRAIRIE AVENUE • CHICAGO, ILLINOIS

L11
E132

EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT

A quarterly journal devoted to the development and application of measures of individual differences.

418675

EDITOR

G. FREDERIC KUDER.....United States Civil Service Commission

ASSOCIATE EDITORS

DOROTHY C. ADKINS.....Social Security Board
FORREST A. KINGSBURY.....University of Chicago
M. W. RICHARDSON.....Adjutant General's Office, A. U. S.

BOARD OF COOPERATING EDITORS

RICHARD D. ALLEN
Providence Public Schools

JOHN G. DARLEY
University of Minnesota

HAROLD A. EDGERTON
Ohio State University

MAX D. ENGELHART
Chicago City Junior Colleges

E. B. GREENE
Social Security Board

J. P. GUILFORD
University of Southern California

E. F. LINDQUIST
State University of Iowa

P. J. RULON
Harvard University

DAVID SEGEL
U. S. Office of Education

C. L. SHARTLE
Social Security Board

H. C. TAYLOR
Western Electric Company

THELMA G. THURSTONE
Chicago Teachers College

HERBERT A. TOOPS
Ohio State University

E. G. WILLIAMSON
University of Minnesota

BEN D. WOOD
Columbia University

The journal is open to (1) reports of research on the development and use of tests and measurements in education, government, and industry, (2) descriptions of testing programs being used for various purposes, (3) discussions of problems of measurement in general or in specific fields, and (4) miscellaneous notes pertinent to the measurement field, such as suggestions of new types of items or improved methods of treating test data. Manuscripts should be sent to EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 1700 Prairie Avenue, Chicago, Illinois.

EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT is published quarterly by Science Research Associates, 1700 Prairie Avenue, Chicago, Illinois. Subscription rate, $4.00 a year. Entered as second class matter June 11, 1941, at the Post Office at Chicago, Illinois, under the Act of March 3, 1879.

EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT

Volume II JANUARY, 1942 Number 1

THE CONCEPT OF OCCUPATIONAL ADJUSTMENT 3
Walter A. Lurie

COMPLETELY WEIGHTED VERSUS UNWEIGHTED SCORING IN AN ACHEIVEMENT EXAMINATION 15
J. P. Guilford, Constance Lovell, and Ruth M. Williams

A PRELIMINARY STUDY OF THE RELATION OF MEASURED INTEREST PATTERNS AND OCCUPATIONAL DISSATISFACTION 23
Theodore R. Sarbin and Hedwin C. Anderson

MEASUREMENT ASPECTS OF THE NATIONAL CLERICAL ABILITY TESTING PROGRAM 37
William J. E. Crissy and M. J. Wantman

INTROVERSION-EXTROVERSION AS A FACTOR IN TEACHER-TRAINING 47
Catharine Evans and C. Gilbert Wrenn

AN INVESTIGATION OF THE POSSIBILITIES OF MEASURING PERSONALITY TRAITS WITH THE STRONG VOCATIONAL INTEREST BLANK 59
Lyle Tussing

A STUDY OF THE GENTRY VOCATIONAL INVENTORY 75
Clifford Froehlich

THE RELATIONSHIP OF THE AFFECTIVE TOLERANCE INVENTORY TO OTHER PERSONALITY INVENTORIES 83
Robert I. Watson

THE INFLUENCE OF TRAINING ON MECHANICAL APTITUDE TEST SCORES 91
Richard W. Faubion, Earle A. Cleveland, and Thomas W. Harrell

NEW TESTS 95

MEASUREMENT ABSTRACTS 99

Copyright, 1942, by
SCIENCE RESEARCH ASSOCIATES

PRINTED IN THE UNITED STATES OF AMERICA

THE CONCEPT OF OCCUPATIONAL ADJUSTMENT[1]

WALTER A. LURIE
Jewish Vocational Service and Employment Center, Chicago

I. THE PROBLEM

THE NEED for criteria of occupational adjustment arises from the attempt to evaluate educational and vocational guidance programs. Many criteria have been proposed, most of them falling into one of the following groups: earnings, job performance, job satisfaction, stability of employment, level of work done, social value of work done, and realization of potentialities.

Several more recent papers have dealt specifically with the weaknesses of studies based on various criteria. Stott (2), in summarizing British experience with a number of the proposed criteria, stressed particularly the sources of unreliability in each of the suggested estimates. Williamson and Bordin (8) have reviewed studies which they are careful to designate as "evaluation of counseling programs" rather than of "adjustment," pointing out structural defects and specific weaknesses in these investigations. Viteles, who had in 1932 suggested accepting "satisfaction and economic efficiency as independent criteria of adjustment in work" (6, p.140), in 1936 proposed a clinical measure, the "dynamic criterion" (7), based upon the extent to which the individual had realized his capacity for vocational success. Williamson and Bordin (8) advocate a "judgment criterion," also a clinical estimate.

These various approaches delimit four possible methods of evaluating vocational adjustment.

[1]I wish to acknowledge my gratitude to Dr. Irving Lorge for supplying me with some of the data used in this study. I also wish to thank Mr. S. T. Friedman for assistance with the computations.

A. The first method is to accept one of the proposed criteria as the essence of occupational adjustment and to define the various degrees of excellence of adjustment in terms of that variable alone. Investigations of vocational adjustment are simply made in terms of the relative job-satisfaction of the individuals, or of their relative earnings, and so forth. The choice of any single criterion is obviously arbitrary.

B. The second possible method of evaluating vocational adjustment is to observe two or more of the criterion variables and to combine the ratings into a single score of vocational adjustment. One might, for instance, weight job satisfaction as 5, skill as 2, earnings as 2, job status as 1, and designate the weighted criterion score as a measure of occupational adjustment. Even if a more complex functional relationship is postulated, the arbitrary and inflexible nature of this procedure is immediately apparent.

C. The clinical or judgmental method of evaluating adjustment is in most current use. Williamson and Bordin (8, p.17) describe the "judgment criterion" as one "by means of which the adjustment of the student is *estimated* in terms of his original problems and of the available data, including the part-criteria" (i.e., the various separate criteria which have been proposed). In contrast to the first two methods, the clinical method uses all the available data, rather than a restricted set. The method of combining data is not mechanical and inflexible, as in an arithmetical combination. The person making the judgment is expected to take individual circumstances into consideration, in an effort to obtain an integrated representation of the adjustment of each unique personality. By the exercise of clinical insight, when combining data, the judges in effect assign a set of weights characteristic of the individuals judged: earnings are unimportant for John Doe, who comes from a wealthy family; the neurotic Jane Smith will never be more satisfied in any other job than in this one; job status is particularly important for Richard Roe because of his brother-in-law's position in the community; and similar examples. While the skill and intuition of the counselor

obviously play an important role in assigning individuals to places on the scale of vocational adjustment, fair agreement is obtained among judges (9).

There is, however, one basic assumption upon which the validity of the clinical method as an instrument of analysis depends entirely. *It assumes that there is such a psychological entity as occupational adjustment,* that it is therefore possible to define a linear continuum corresponding to excellence of adjustment in various degrees and to locate individuals at definite points on this scale. The clinical method of evaluation shares this fundamental assumption with the first two procedures, the selection of a single criterion and the combination of criteria by formula. It differs only in advocating clinical judgment as the instrument for assigning individuals to points on the scale of adjustment.

It is not my purpose, in questioning the validity of this assumption, to disparage the role of clinical insight in counseling. No one who has attempted to guide individuals in the choice of careers and preparation for them, and later to evaluate the results, can deny that clinical intuition gives more full-bodied and meaningful results than the use of a single criterion or the mechanical combination of criteria. This does not preclude the possibility that the method rests upon a faulty premise. It is always possible to project data of any degree of complexity upon a single axis, either by formula or flexibly, on a case-by-case basis. But this process will necessarily be arbitrary and meaningless if the structure is not, in actual fact, unidimensional. If there is no such linear continuum as occupational adjustment, all three methods which we have considered would be invalidated, the clinical as well as the other two. The superior satisfaction which the counselor receives from the use of the clinical method may reflect an actual lack of precision, which covers up more effectively than the mechanical procedure the incompatibility of data combined in his judgment. The only check upon his estimates, namely, his agreement with other judges, may show merely the extent to which they share his preconceptions.

D. It is therefore advisable to seek a fourth procedure for evaluating occupational adjustment, differing from the first three in its fundamental premise. The nature of such a procedure is clear in a simpler, but fully analogous, situation.

Let us suppose that it is our problem to evaluate the size of human adults, instead of their occupational adjustment. The simplest thing to do would be to define size in terms of, let us say, weight. This would be comparable to identifying occupational adjustment with job satisfaction. But why not choose height, or volume? The choice is arbitrary. A second possibility is to develop a formula combining height and weight. This would require many individual exceptions, because of differences in sex, age, skeletal structure, incidence of crippling accident or disease, and other factors. Would the next proposal then be to substitute a clinical or judgmental combination of all the various factors for the purpose of arranging individuals in order of size? It is more likely that investigators would suspect size to be a variable which cannot be evaluated as such, because no continuum corresponds to the concept. Efforts would be directed first towards defining a set of variables which have something to do with size as popularly conceived and which can be observed and predicted. A further step would be to investigate the dimensionality and structure of the "size" variables. Basic variables—primary factors—would be identified in terms of which all the "part-criteria" of size could be studied in relation to various hereditary and environmental influences.

Now let us reverse the analogy. It is the contention of this paper that occupational adjustment, like size, is a composite variable. Since many part-criteria have already been defined, the next step in evaluating vocational adjustment as popularly conceived is to investigate its dimensionality and structure. Unless this next step is taken, we must work with a concept of occupational adjustment which brings together in a single rating or judgment different factors which have meaning separately but which cancel each other when an effort is made to combine them by formula or by the exercise of clinical insight.

II. THE EVIDENCE

The choice among the four proposed methods can be submitted to consideration in the light of evidence. If the evidence shows that there is a single factor common to all proposed criteria which can in themselves be considered meaningful, the concept of vocational adjustment as a psychological entity will have been upheld. Whether we should then use a single criterion, a combination by formula, or a clinical estimate would be a practical problem, depending upon which could be shown to arrange individuals most accurately in order of excellence of vocational adjustment. If, however, several independent factors are demonstrated, it would obviously be wiser to regard these as separate criterion-variables, which must be observed separately and thought of separately as goals of guidance programs. A crucial test of these various approaches can, therefore, be applied by the factor analysis of criterion data.

Table 1 B shows the intercorrelations of five criterion variables (Table 1 A) used in Thorndike's (3) study of prediction of vocational success, for 175 men in the biennium 24-25. Table 1 C shows the distribution of tetrad differences from these correlations. It seems likely, in view of the large deviations from zero, that one factor is not sufficient to account for the intercorrelations of these carefully collected data.

Table 2 B shows the tetrachoric correlation coefficients (1) of twelve criterion and background items (Table 2 A) for 55 female job-applicants born in 1914, whose contact with the Jewish Vocational Service occurred in 1937. Table 2 C gives the centroid factor loadings (4), Table 2 D the distribution of residuals after extraction of three factors, Table 2 E the final factor loadings, and Table 2 F the matrix of transformation by which the centroid factor loadings in Table 2 C were rotated to obtain the final factor loadings as given in Table 2 E (5). Table 2 G shows that the factors are by no means identical: the greatest deviation from orthogonality is less

than 25 degrees, and one plane is orthogonal to both of the others. The factors can be identified only tentatively, in view of the few subjects and the difficulty in obtaining precise information about the clients of an employment agency. Table 2 H lists the high and zero loadings for each factor. Factor I seems to be a reflection of the amount of work experience, which was not controlled; Factor II is primarily a matter of job-satisfaction or job-level; and Factor III is an employability factor. At any rate, it is clear that the criterion vectors do not lie along a single axis.

III. CONCLUSIONS AND DISCUSSION

A. It is my belief that the evidence warrants a tentative verdict in favor of the fourth approach, that of discarding the concept of occupational adjustment as a psychological entity and observing separately instead several dimensions of occupational adjustment. While the force of this study may be diminished because it was necessary to use weak criteria and scanty data, and in particular because no clinical estimates could be included, it is certainly not invalidated by these faults. The burden of proof rests now upon those who would evaluate occupational adjustment on a linear scale. They must show that the polydimensionality demonstrated in this study reflects the introduction of irrelevancies and that all meaningful criteria *can* be projected on a linear scale.

B. The present study is not sufficiently comprehensive to identify clearly the dimensions of vocational adjustment. It seems likely, however, that job-satisfaction is closely identified with one factor, and ease of obtaining employment with another. Fruitful investigations into the nature of vocational adjustment can be directed towards further clarification of these factors and of their relationship to the potentialities, background, and aspirations of the individuals.

C. These results, even if they are accepted as conclusive, do not by any means throw into question the efficacy of clinical insight in the counseling process. Recommendations to individuals regarding their life-conduct can be made only by

skilled advisers, never automatically by formula. The clinical method may have a place in evaluation as well as counseling; job satisfaction, for instance, may be better estimated than measured by objective techniques, and this may be true of any other variable which is a psychological entity. Perhaps, however, if these findings are correct, counselors will be able to sharpen their conception of the goals which they are attempting to promote for each individual. When they evaluate informally the results of counseling in a particular case, they may find three or four sentences more informative than an over-all judgment regarding the excellence of occupational adjustment.

D. Finally, it cannot be argued that we are justified in projecting all criterion data on a single axis because it is *important* to have an over-all evaluation of adjustment. What is important is to know the nature, the character of each individual's vocational adjustment. We must study the whole person in action in terms of meaningful variables of the occupational adjustment complex, but we need not place him on a meaningless scale.

IV. SUMMARY

A. It was shown, by logical analysis of various possible methods of evaluating excellence of occupational adjustment, including clinical judgments, that they all assume such estimates or judgments to form a linear continuum.

B. Evidence was presented that only polydimensional models can represent adequately two sets of data, one from Thorndike's major study of vocational success, one derived from clients of the Jewish Vocational Service.

C. Without questioning the efficacy of clinical insight in the guidance process, it was suggested that the concept of occupational adjustment as a psychological entity should be supplanted by a concept of occupational adjustment as a complex of factors which must be observed separately and can be considered separately as goals of guidance programs.

D. Further studies should be directed towards the more precise identification of these factors. This first tentative

identification would suggest that one has to do with job-satisfaction and one with ease of obtaining employment, in addition to others not as yet identified.

TABLE 1*

Data from Thorndike's study

175 men in biennium 1924-1925

1 A. *Variables*

1. Earnings
2. Level of job
3. Interest in job
4. Time unemployed
5. Change of employer

1 B. *Intercorrelations***

Var.	1	2	3	4	5
1		45	21	–60	–34
2			32	–14	–18
3				–06	–15
4					40
5					

The standard errors of these coefficients range from .03 to .07.

1 C. *Tetrad differences****

Absolute Value	Frequency
00	1
01	2
02	0
03	1
04	1
05	0
06	2
07	3
08	0
09	0
10	0
11	1
12	1
13	1
14	0
15	0
16	2

*Decimal points have been omitted.

**I wish to thank Dr. Irving Lorge for supplying me with this table of intercorrelations.

***Only one of each pair of tetrad differences identical except for sign has been recorded.

TABLE 2*

Data from JVS & EC clients

55 women born in 1914, data as of 1937

2 A. *Variables*

1. Time of leaving full-time school; 1933 or later recorded as minus (41% recorded as minus)
2. Number of months employed to 1937; 39 or less recorded as minus (43%)
3. Minimum weekly wages on jobs reported; $12 or less, minus (38%)
4. Maximum weekly wages reported; $15 or less, minus (42%)
5. Number of different employers reported; 2 or less, minus (46%)
6. Minimum wage stated to be acceptable; less than $15, minus (29%)
7. Satisfaction with wages; will not consider wage lower than previous maximum, minus (35%)
8. Waiting time for first job; unemployed until year following time of leaving school, minus (39%)
9. Success of JVS efforts to place in job; no placement, minus (63%)
10. College or specialized training beyond 4 year high school; none, minus (37%)
11. Satisfaction with type of work done; seeking other type of work, minus (13%)
12. Freedom from recorded handicap, including speech defect, language handicap, deformity, etc.; handicap present, minus (13%)

*Decimal points have been omitted.

418675

TABLE 2 (Continued)

2 B. *Tetrachoric correlation coefficients*

Var.	1	2	3	4	5	6	7	8	9	10	11	12
1		69	–13	34	00	48	–12	–52	–07	–57	25	11
2			56	62	–18	58	08	00	–07	07	64	18
3				34	–16	35	–24	–12	–23	26	59	19
4					01	59	63	07	–10	13	85	–09
5						21	07	43	32	19	12	80
6							–41	16	00	–06	12	57
7								40	09	20	70	–60
8									20	31	06	65
9										40	–50	10
10											–10	–50
11												47
12												

The standard errors of these correlations range up to about .15.

2 C. *Centroid factor loadings*

Var.	I	II	III	Communality
1	66	–22	–18	52
2	73	36	14	68
3	53	35	32	50
4	49	60	38	75
5	–18	54	–60	68
6	63	33	–43	69
7	–25	42	57	56
8	–30	72	–31	71
9	–43	22	–27	30
10	–44	49	37	57
11	62	62	49	1.00
12	32	43	–86	1.02

2 D. *Distribution of absolute values of third factor residuals*

Absolute Value	Frequency
00-04	11
05-09	13
10-14	10
15-19	8
20-24	13
25-29	4
30-34	5
35-39	0
40-44	2

2 E. *Final factor loadings*

Var.	I	II	III
1	53	–22	–10
2	82	45	00
3	63	52	–11
4	69	76	00
5	01	07	82
6	69	08	44
7	–06	64	–14
8	–02	38	73
9	–33	–02	41
10	–22	57	09
11	82	85	–09
12	43	–11	88

2 F. *Matrix of transformation* (Λ_3)

	I	II	III
I	930	100	–156
II	368	812	618
III	031	576	–770

2 G. *Direction cosines of normals* ($\Lambda'_3\Lambda_3$)

	I	II	III
I		410	058
II			043
III			

TABLE 2 (Continued)

2 H. *Tentative identification of factors*

Factor I.

A. *Variables with loadings over 40*

2.	Number of months employed	82
11.	Satisfaction with type of work	82
6.	Minimum wage acceptable	69
4.	Maximum weekly wages	69
3.	Minimum weekly wages	63
1.	Time of leaving school	53
12.	Freedom from handicap	43

B. *Variables with loadings under 15*

5.	Number of different employers	01
8.	Waiting time for first job	02
7.	Satisfaction with wages	–06

C. *Tentative identification*

Experience

Factor II.

A. *Variables with loadings over 40*

11.	Satisfaction with type of work	85
4.	Maximum weekly wages	76
7.	Satisfaction with wages	64
10.	Special training	57
3.	Minimum weekly wages	52
2.	Number of months employed	45

B. *Variables with loadings under 15*

9.	JVS placements	–02
5.	Number of different employers	07
6.	Minimum wage acceptable	08
12.	Freedom from handicap	–11

C. *Tentative identification*

Job level or job satisfaction

Factor III.

A. *Variables with loadings over 40*

12.	Freedom from handicap	88
5.	Number of different employers	82
8.	Waiting time for first job	73
6.	Minimum wage acceptable	44
9.	JVS placements	41

B. *Variables with loadings under 15*

2.	Number of months employed	00
4.	Maximum weekly wages	00
10.	Special training	09
11.	Satisfaction with type of work	–09
1.	Time of leaving school	–10
3.	Minimum weekly wages	–11
7.	Satisfaction with wages	–14

C. *Tentative identification*

Ease in finding employment

REFERENCES

1. Chesire, L., Saffir, M., and Thurstone, L. L. *Computing Diagrams for the Tetrachoric Correlation Coefficient.* Chicago: University of Chicago Press, 1933.

2. Stott, M. B. "Occupational Success." *Occupational Psychology* (London), XIII (1939). 126-140.

3. Thorndike, E. L., *et al. Prediction of Vocational Success.* New York: The Commonwealth Fund, 1934.

4. Thurstone, L. L. *The Vectors of Mind.* Chicago: University of Chicago Press, 1935.

5. Thurstone, L. L. "A New Rotational Method in Factor Analysis." *Psychometrika,* III (1938). 199-218.

6. Viteles, M. S. *Industrial Psychology.* New York: W. W. Norton and Co., 1932.

7. Viteles, M. S. "A Dynamic Criterion." *Occupations,* XIV (1936), 962-967.

8. Williamson, E. G. and Bordin, E. S. "The Evaluation of Vocational and Educational Counseling: A Critique of the Methodology of Experiments." *Educational and Psychological Measurement,* I (1941), 5-24.

9. Williamson, E. G. and Bordin, E. S. "A Statistical Evaluation of Clinical Counseling." *Educational and Psychological Measurement,* I (1941), 117-132.

COMPLETELY WEIGHTED VERSUS UNWEIGHTED SCORING IN AN ACHIEVEMENT EXAMINATION

J. P. GUILFORD, CONSTANCE LOVELL, and RUTH M. WILLIAMS
University of Southern California

IN A PREVIOUS REPORT,[1] the senior author presented the derivation of a scoring weight for differential weighting of responses to test items. The formula for the weight, in a form recommended for practice, is as follows:

$$W = 4 + \frac{p_u - p_l}{\sqrt{pq}},$$

in which W = the scoring weight,

p_u = the proportion of an upper (or otherwise specified criterion sub-group) reacting in a defined manner,

p_l = the similar proportion in a lower (different) criterion sub-group,

p = the proportion of the two sub-groups combined responding in this manner, and

$q = 1 - p$.

Such a weight has heretofore been most widely employed in connection with personality tests of the type of the *Strong Vocational Interest Blanks*. In the following study the weight was used in connection with an objectively-scored multiple-choice achievement examination. In this kind of test we can consider the probability of a specified response being made *(p)* or not being made *(q)*, for a group as a whole, and

[1] J. P. Guilford. "A Simple Scoring Weight for Test Items and Its Reliability," *Psychometrika*, VI (1941), 367-374.

also the probabilities of the same response being made within two separated groups. Our main problem was to determine whether an examination with completely weighted scoring of this kind yields any more highly reliable and valid scores than the same examination yields with unweighted scoring. A subsidiary problem was to determine whether the length of examination has any bearing upon the effect of weighted versus unweighted scoring. By "completely weighted" we mean that *every* response, whether considered right or wrong, is given a weight in proportion to its predictive significance. This procedure is in contrast to ordinary differential weighting where only correct responses are weighted in proportion to their diagnostic value. By "unweighted scoring" we mean that items are given weights of 0 or 1 — 0 if a wrong response of any kind is given and 1 if the correct response is given.

As our experimental material we used the results of a final examination in a course on "Problems of Human Behavior."[2] The test was composed of 201 multiple-choice items and 107 true-false items. These items had been analyzed for validity in previous use; therefore, we could expect to find an unusually large number of diagnostic items both when scoring was weighted and when it was not. Instructions to guess or not to guess were not stated on the test blanks, but the usual correction formula had been applied to allow for guessing. There were extremely few omissions. These facts are mentioned because we used the total examination score as our criterion of achievement in the course.

Our study was confined to the first 100 consecutive items in the examination, skipping one item which had an unusual number of omissions and which had been excluded from the scoring of the total examination. All but 8 of the items proved to have phi coefficients of .14 or larger (in other words, significant correlations)[3] and all but 19 had very significant phi

[2]We are indebted to Dr. Neil Warren for the opportunity to use this material.

[3]J. P. Guilford. "The Phi Coefficient and Chi Square as Indices of Item Validity," *Psychometrika*, VI (1941), 11-19.

coefficients (greater than .18). The total range of phi coefficients for the 100 items was from −.06 to +.48, with a median of + .28.

Three hundred test papers, selected at random, were used in this investigation. The papers were re-scored in order to be absolutely sure about total, or criterion, scores. The upper and lower criterion sub-groups were composed of the 100 highest and the 100 lowest ranking students in the list of 300. The proportion of each group responding in each of the four ways to every item was determined. The scoring weight for each particular response was read from a graphic chart.[4] For example, one item read: "Genius and feeble-mindedness are (1) points on a normal distribution curve; (2) points on a bimodal distribution; (3) points on a multimodal distribution; (4) in separate distributions." The scoring weights for responses 1 to 4 inclusive were: 6, 3, 4, and 2, respectively.

A logical reason for expecting improved reliability and validity from weighted scoring is now more apparent. Not all the wrong responses are of equal diagnostic value, an assumption that is implicitly made in unweighted scoring. It is apparently worse for a student to err by choosing answer 4 and less serious for him to choose answer 3. In only 15 items out of the 100 did the three wrong answers prove to have equal weight. No weights exceeded 6 points nor fell below 2 points on a scale that extended from 0 to 8. This range was to be expected from the moderate and small sizes of the phi coefficients found for correct responses, as previously mentioned.

The factor of length of test was investigated roughly by selecting two shorter examinations composed of the first 20 and the first 50 items out of the 100. Each of the three tests of different length was scored in both halves (odds and evens) and in total, with and without scoring weights. For this purpose, 100 papers were selected out of the original list of 300, by taking every third paper when the 300 were in rank order for total scores. The correlations to be mentioned next are based upon these 100 papers.

[4]See reference in footnote 1.

TABLE 1

RELIABILITY AND VALIDITY COEFFICIENTS

Length of Test	Reliability		Validity	
	Weighted Scoring	Unweighted Scoring	Weighted Scoring	Unweighted Scoring
20 items	.667	.649	.817	.793
50 items	.860	.844	.892	.901
100 items	.922	.899	.900	.924

The reliability coefficients were estimated by the Spearman-Brown formula in each case. The correlation of each short test with the total (criterion) score was computed. The reliability and validity coefficients are summarized in Table 1. Here it is obvious that the weighted scoring yielded a scant average gain of .02 in the reliability coefficients. This trifling gain is consistent, but seems to be insignificant. In validity the weighted scoring yielded a gain of about .02 in the shortest test and a like amount of loss in the longest test, neither of these changes being significant.

It is well to consider possible special reasons for the failure to obtain increased reliability and validity here. As indicated above, the phi coefficients between items and criterion scores were generally low and the range of differential weights was relatively small. It might be that in other examinations which include items with weights extending closer to 0 and 8 there would be an appreciable gain from differential weighting. On the other hand, 32 of our 100 items had weights ranging from 2 to 6; 21 more had weights ranging from 2 to 5 or from 3 to 6; and 45 had weights ranging from 3 to 5, to be contrasted with an unweighted range of 0 to 1.

One important source of gain to be expected from complete differential weighting comes from the variations in weights among the wrong responses, which in unweighted scoring or in partially weighted scoring are all given the same value of zero. Gaps of more than one point between the correct response on the one hand and all the wrong responses on the other simply magnify numerically the variability among individuals' total scores, except that the more diagnostic items are then allowed to contribute relatively more to the total

variability than do the less diagnostic ones. When there is a spread of the weights among the wrong responses, there should be more refinement in providing effectual variability among total scores. The weights for wrong responses among our 100 items showed very narrow ranges. 15 items had equal weights for wrong responses, 68 had differences not exceeding one, and 17 had differences not exceeding two points. While these differences are small, it would seem that their effects should have been felt in scoring.

Another possible factor in the failure of weighting is that the criterion scores were derived from unweighted scoring. Correlations of either weighted or unweighted scores for the shorter tests of 20, 50, and 100 items and total scores are in a sense spurious in that we are correlating part with whole. This factor might have favored higher correlations, especially for the unweighted scoring of the parts.

In using total scores as our criterion of achievement here, we have assumed up to this point that, though the part-whole correlations are spurious, they are equally so for both types of scoring. One evidence that this assumption may not be sound is the fact that the shorter the part-test, the relatively greater is the advantage of weighted over unweighted scoring. (See Table 1) On the other hand, it may be characteristic of short tests *per se* to gain relatively more from weighted scoring. Other evidence is found among the reliability coefficients. Here we have unweighted scoring correlated with unweighted scoring and weighted scoring correlated with weighted scoring. The correlations indicate that regardless of the length of the test within the range of 20 to 100 items, there is about the same slight advantage for weighted scoring. Although changes in validity do not always parallel changes in reliability, it would seem that if the evidence of reliability here is dependable, the systematic variations in validity may be due to the relatively greater spuriousness for unweighted scoring in the longer tests (since they are greater parts of the total and are similarly scored) rather than to any greater relative gain in validity of short tests by weighted scoring. The evidence is

entirely too meager, however, for us to draw any final conclusions on this point.

In view of the uncertainty introduced by the factor of spuriousness of correlation, it would have been interesting to see what would have happened with an outside criterion. Some idea of the extent of spuriousness can be obtained, without taking the trouble to score the tests minus the 20, or 50, or 100 items used in the experimental tests, by applying a formula to estimate the amount of correlation between part and the whole from which the effects of the part are eliminated.[5] For the 50-item test, with unweighted scoring, for example, this estimated correlation is .862, which may be taken as an indication of the correlation between the 50-item test and an outside criterion of about 250 items. The correlation of the same test with a homogeneous test of 300 items (the approximate length of the total examination) is estimated by formula to be .865.[6] The amount of spuriousness is then indicated by the difference between .865 and .901. We cannot similarly estimate the amount of spuriousness in the correlation of the weighted scoring in the 50-item test since it is not a simple part-whole relationship. But had the spuriousness in this case been zero, the validity coefficient of .892 is not quite .03 higher than that of the estimated unweighted scoring without its spurious element. It is doubtful, therefore, whether the hypothesis of greater part-whole spuriousness attributed to the unweighted scoring is sufficient to account for the failure of weighted scoring to exhibit superior validity coefficients.

Had all the items in our tests been significantly correlated with the criterion, a difference in favor of weighted scoring might have resulted. Therefore, we selected for comparison the 50 items of highest diagnostic value. The reliability coefficients were then .874 and .873 for weighted and unweighted scoring, respectively, and the corresponding validity coefficients

[5]C. C. Peters and W. R. Van Voorhis, *Statistical Procedures and Their Mathematical Bases* (New York: McGraw-Hill Book Company, 1940), p. 217.

[6]J. P. Guilford, *Psychometric Methods* (New York: McGraw Hill Book Company, 1936), p. 422.

were .900 and .904. These coefficients were insignificantly better than the ones derived from the 50-item test in which the items were taken at random. The inference might be that the items used in all our experimental tests were at a sufficiently high level of diagnostic value, taking them collectively, that weighted scoring was of no consequence.

Our general conclusion is that our logically defensible system of completely weighted scoring did not yield an appreciable gain in either reliability or validity in achievement examinations of from 20 to 100 items. While the result is negative so far as the improvement of test technique is concerned, it is useful to know that the customary unweighted scoring, which takes distinctly less time and effort, gives about as reliable and valid results as differential weights afford.[7] Although this result may not be generalized to all weighting methods and to all kinds of tests, it does suggest the possibility of satisfactory scoring without weighting in places where we now attempt to extract the utmost validity by the use of differential weights. With the increased use of machine scoring, where differential weighting becomes a serious practical problem, it may be well in any case to consider the efficiency of weights 0 and 1 before recommending a system of differential scoring weights.

[7]This general outcome is in line with a conclusion reached on rational grounds by M. W. Richardson, in Paul Horst's *The Prediction of Personal Adjustment* (New York: Social Science Research Council, 1941). 379-401.

A PRELIMINARY STUDY OF THE RELATION OF MEASURED INTEREST PATTERNS AND OCCUPATIONAL DISSATISFACTION[1]

THEODORE R. SAKBIN
University of Minnesota

and

HEDWIN C. ANDERSON
Minnesota Division of Vocational Rehabilitation

THAT occupational dissatisfaction is associated with a lack of interest typical of successful men in a particular job is a generally accepted hypothesis. This is of special concern to psychologists, particularly if the hypothesis can be verified and predictions of occupational satisfaction made on the basis of interest measurement. In order to test the hypothesis two kinds of data must be analyzed: (1) evidence of occupational dissatisfaction and (2) measures of vocational interest.

Although rating scales for determining job satisfaction have been developed by Hoppock (4) and others, they are difficult to use in a clinic where the clients or patients form a heterogeneous population. They come from many different walks of life, and there are seldom more than a few individuals who are employed by the same organization. Job satisfaction can be described by Hoppock's definition as "any combination of psychological, physiological or environmental circumstances that causes a person truthfully to say 'I am satisfied with my job.'" (4:47)

[1]This study is one of a series of studies in process on clinical problems of interest measurement at the University of Minnesota Testing Bureau.

In commercial and industrial organizations, a psychologist may experience difficulty in persuading workers "truthfully" to state their feelings about their work because they are afraid of losing their jobs. During periods of widespread unemployment, especially, an individual may express satisfaction with his job merely because it is a job. By dealing with groups of workers and guaranteeing anonymity by the use of unsigned questionnaires, a psychologist may gather group data, but the anonymity may prevent his relating these data to such variables as personality traits or interests in subsequent clinical study.

The clinical situation in which the present data were gathered gives greater assurance of meeting Hoppock's qualification regarding the truthfulness of the clients' responses. In the first place, all the subjects came to the University of Minnesota Testing Bureau voluntarily. They had heard of the Bureau through friends or business associates. They recognized the Bureau as a disinterested organization which each year assists a small number of out-of-school adults with problems of vocational adjustment. Secondly, they paid a special fee for the service. This fact presumably predisposed them to tell the truth about their occupational experiences. Finally, if a client had difficulty in expressing himself, a trained clinical interviewer assisted him to say the things that he could not or would not otherwise have said. It is reasonable to assume from these three facts that expressions of occupational dissatisfaction were truthful expressions. Having found a usable index of occupational satisfaction, the next step was to find a measure of vocational interest.

According to a recent poll (1), the most widely used measure of vocational interest is the *Strong Vocational Interest Blank*. This instrument is based upon this fundamental assumption:

"If a man likes to do the things which men like who are successful in a given occupation and dislikes to do the things which these same men dislike to do, he will feel at home in that occupational environment. Seemingly, also, he should be

more effective there than somewhere else because he will be engaged, in the main, in work he likes." (6)

The *Strong Vocational Interest Blank* was standardized upon people who were purportedly successful in their occupations. Strong's criteria of occupational success include the following: length of experience in an occupation, annual income, level of education, certification of membership in professional society, and selection by so-called competent authorities. These were used singly or in various combinations.

This is not the place to list the description of Strong's criterion groups but as an illustration we take three occupations. The samples of successful men have all been engaged in the respective occupations for at least the three previous years, and none is over 60 years of age.

Accountant: "Includes 160 general accountants, 54 cost accountants, 65 auditors, and 66 comptrollers and treasurers. Average age equals 37.4 years; education equals 12.3 grade."

Office Worker: "Includes 214 office clerks, bookkeepers and stenographers; 92 office managers; and 200 credit managers. Average age equals 33.2 years; education equals 11.5 grade."

Physician: "Graduates of Yale and Stanford Medical School. Includes 252 physicians and 75 surgeons (no difference of interest between them). 253 are from California, 47 from Connecticut and 9 from New York; the remaining are scattered. Average age equals 40.9 years; education equals 18.5 grade."

In interpreting the results on the *Strong Vocational Interest Blank,* then, it is always necessary to think of the criterion groups which served as the norm, as well as the percentages of the group included under each grade. Thus if an individual scored A on the key for physicians it means that he made a score in the range of the top 69 per cent of the physicians who made up the norm group. A score of B falls in the range of the next 29 per cent; and a score of C falls in the range of the lowest two per cent of the criterion group (3).

For purposes of the present analysis, 100 cases were

selected from the files of the University of Minnesota Testing Bureau for the period 1937 to 1940 on the basis of completeness of data. This sample contained 76 men and 24 women. The cases were so-called "non-college adults", individuals who are accepted by the University Testing Bureau for research and clinical purposes. Those who had gross physical abnormalities, such as paralysis, spasticity, and deafness were not included in this selection of cases. Only individuals who were 25 years of age or more were included. Above this limiting age individuals usually have had some opportunity to establish a work history. The mean age for men was 31.5 with a standard deviation of 5.7 and a range of 25-53; for women, 30.4 with a standard deviation of 7.7 and a range of 25-44.

The educational level of this group appears to be higher than that of the general population. For men, the mean grade completed was 13.8, S.D. 2.5; for women, 14.5, S.D. 2.3.

The occupational status of this group was also higher than that of the general population. According to the *Minnesota Occupational Rating Scales,* 72 per cent rated in the top three categories. In the general population only 22 per cent fall into these three categories.[2]

From an analysis of these data, the following hypothesis can be tested:

Adults who express dissatisfaction with their current[3] occupations show no primary pattern of interest, as measured by the *Strong Vocational Interest Blank,* for the group of occupations in which their current occupation belongs.

The *Strong Vocational Interest Blank* was first analyzed in order to determine the primary pattern of interest. Darley's

---

[2]Occupational Class I: professional; Occupational Class II: semi-professional and managerial; Occupational Class III: clerical, skilled trades, retail business. F. L. Goodenough and J. E. Anderson, *Experimental Child Psychology,* (New York: Appleton-Century, 1931), pp. 501-12.

[3]The data contained herein are concerned with present occupation (or most recent occupation for four cases unemployed at the time of counseling). Data were available on *modal* occupations but since these coincided with present occupations in over 90 per cent of the cases, the data were not analyzed in terms of the modal occupations.

scheme of determining the presence and intensity of patterns of interest on the Strong Blank was utilized (3). A primary pattern is defined as a preponderance of A and B+ scores within the occupations making up a group of factors as revealed by existing factor analysis studies. To illustrate: the verbal or linguistic interest type, Group X on the Strong test, is made up of the following typical occupational titles—advertising man, editor, lawyer. If a client had scores of A, B+, A respectively, on these three keys, he would be considered to have a primary pattern of interest in this group of occupations. If his scores were B+, B, B, he would be rated as having a secondary pattern of interests. A tertiary pattern of interests is defined as a majority of B and B— scores on the keys within any factor or group. In the present study the number of cases was too small to be treated in terms of Darley's fourfold classification: primary, secondary, tertiary, and no pattern. Instead, we considered only two categories:

(a) Presence of primary pattern (this means presence of primary pattern as defined above in the group which embraces the client's present or most recent occupation).
e.g. Client's present occupation: Automobile Salesman;
Scores on Strong Blank Group IX

Real Estate Salesman	B+
Life Insurance Salesman	A
Sales Manager	A

(b) Absence of primary pattern (this means absence of primary pattern in the group which embraces the client's present or most recent occupation).
e.g. Client's present occupation: Lawyer;
Scores on Strong Blank Group X

Advertising man	B—
Lawyer	C
Editor	C

Each case was classified according to the client's stated complaint. The following categories were used:

(a) Dissatisfied with occupational field.
(b) Dissatisfied with present job.
(c) Dissatisfied with present job only because of future prospects.
(d) No specifically stated dissatisfaction, but seeks vocational and/or educational information or advice.

Each case was also classified according to the clinician's diagnosis, using five broad classifications:

(a) *Inappropriate vocational choice:* e.g., "He does not have the interests of salesmen." "Has never been interested in mechanical work." "Working with teachers not congenial to this man's values, attitudes, and ideals."
(b) *Primary personality disorders:* e.g., "social maladjustment", "unhappily married", "neurotic tendencies."
(c) *Insufficient education or training:* e.g., "lacks stenographic skills to compete with co-workers", "lacks sufficient graduate training to get into junior college teaching", "lacks skills in cost estimating which are required for promotion and increased pay."
(d) *Inappropriate job placement:* e.g., "clerical skills not being used", "stenographic skills are not adequate", "truck-driving satisfactory, but would be happier if he had a run closer to home and family", "selling satisfactory, but his product is inappropriate."
(e) *Other:* This includes a small number of which three were characterized as "no problem", the rest as financial, health, or unclassified.

These diagnostic illustrations are stated as single entities. This is somewhat misleading. For many of the cases a multiple diagnosis was made. For example, 28 per cent of the cases of *inappropriate vocational choice* also exhibited neurotic symptoms and mild personality disorders. The data, however, were treated in terms of the diagnosis that was considered by the clinician to be the most significant one.

A third kind of classification was made to determine the

frequency of types of treatment or recommendations. These fall under the following headings:

(a) *Placement advice:* e.g., "Since your interests and abilities fit the picture of successful salespeople, I would recommend that you register with the X and Y employment agencies." "You should seek employment in a more technical field than your present occupation." "You are faced with two alternatives: taking over your father's business, or continuing as an engineer. Your interests and personality traits would suggest that you would be happier as an engineer than as a business man."

(b) *Additional training recommended:* e.g., "A University Extension Course in Cost Estimating seems indicated." "In order to prepare for the position in mind, you will have to return to college for two years of graduate work." "In order to capitalize on your assets and interest, you should obtain the necessary skills at such a school as The Blank Industrial Training Institute."

(c) *Psychotherapy:* e.g., recreational therapy, catharsis, helping client to gain insight into family or other conflict situation, suggestive therapy, group therapy, relationship therapy, and so on.

(d) *Referral to psychiatrist:* Obvious psychiatric problems.

(e) *No advice or recommendations.*

The results of the analyses are summarized in the three tables. Table 1 shows the clinicians' diagnoses for 76 male clients and how they are related to the presence or absence of primary patterns of interest and also to clients' complaints. Table 2 shows the same data for 24 female clients. Table 3 summarizes the treatment techniques as related to diagnoses for the 100 cases.

Table 1 reveals one fact quite clearly: most adult males who complain of occupational dissatisfaction show no primary pattern of interest in the group of occupations which embraces their present occupation. Sixty-two of the 76 men (82 per

TABLE 1

CLIENTS' STATEMENT OF PROBLEM AND CLINICIANS' DIAGNOSES IN TERMS OF PRESENCE OR ABSENCE OF PRIMARY PATTERN OF INTEREST IN CURRENT OCCUPATION

(N=76 Men)

Client's Statement \ Clinician's Diagnosis	Inappropriate vocational choice		Primary personality disorders		Insufficient training		Inappropriate job placement		Other		Totals	
	PRIMARY PATTERN		PRIMARY PATTERN		PRIMARY PATTERN		PRIMARY PATTERN		PRIMARY PATTERN		PRIMARY PATTERN	
	Present	Absent	Present	Absent	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Dissatisfaction with ocupational field	0	18	3	6	0	1	0	1	0	1	3	27
Dissatisfaction with specific job	0	3	3	0	1	0	0	1	0	0	4	4
Dissatisfaction with future of job	0	0	0	1	1	0	0	3	0	0	1	4
No specifically stated dissatisfaction	2	12	2	7	0	2	1	0	1	6	6	27
Totals	2	33	8	14	2	3	1	5	1	7	14	62

TABLE 2

CLIENTS' STATEMENT OF PROBLEM AND CLINICIANS' DIAGNOSES IN TERMS OF PRESENCE OR ABSENCE OF PRIMARY PATTERN OF INTEREST IN CURRENT OCCUPATION

(N=24 Women)

Client's Statement \ Clinician's Diagnosis	Inappropriate vocational choice		Primary personality disorders		Inappropriate job placement		Other		Totals	
	PRIMARY PATTERN		PRIMARY PATTERN		PRIMARY PATTERN		PRIMARY PATTERN		PRIMARY PATTERN	
	Present	Absent	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Dissatisfaction with occupational field	0	4	1	0	0	0	0	0	1	4
Dissatisfaction with specific job	0	1	0	1	1	1	0	1	1	4
No specifically stated dissatisfaction	1	0	3	4	0	1	4	1	8	6
Totals	1	5	4	5	1	2	4	2	10	14

TABLE 3

ANALYSIS OF TREATMENT TECHNIQUES IN TERMS OF CLINICIAN'S DIAGNOSIS

(N=76 Men, 24 Women)

Treatment Used \ Clinician's Diagnosis	Inappropriate vocational choice		Primary personality disorders		Insufficient education		Inappropriate job placement		Other		Totals	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Placement advice	11	2	4	3	1	0	4	3	3	0	23	8
Recommended additional training	19	4	3	0	4	0	2	0	1	2	29	6
Psychotherapy	3	0	11	3	0	0	0	0	0	2	14	5
Referral to psychiatrist	0	0	3	3	0	0	0	0	1	0	4	3
No advice	2	0	1	0	0	0	0	0	3	2	6	2
Totals	35	6	22	9	5	0	6	3	8	6	76	24

cent) fall into this category. When we consider the women who came to the Testing Bureau, the association is not so clear cut. (See Table 2) Fourteen of the 24 women clients (58 per cent), had no primary pattern in the occupation in which they were employed. When we consider the 10 women who actually *expressed* dissatisfaction with their work (items 1 and 2 in Table 2), we see that eight (80 per cent) had no primary pattern in the group of occupations which embraced their present employment. These equivocal results in the case of the women may be attributed to the inadequacy of the Strong Blank for Women, to the smaller number of cases, or to the generally accepted statement of sex differences in intensity and variety of interests.

Where the men actually expressed dissatisfaction, 70 per cent referred to the occupational field rather than to the particular job in which they were employed at the time of the interview (Items 1 and 3 versus totals of items 1-2-3). For example, one junior high school teacher said: "It isn't my job at the Blank Junior High School that I don't like. As teaching jobs go, it's a good one. It's the idea of spending the rest of my life as a teacher that is my bogey." A small number did express dissatisfaction with their present jobs, but considered the occupational field in which the job was located as a desirable one. Only five individuals expressed satisfaction with the field of the occupation and with the present job, but were concerned over the future prospects of the job.

Analysis of the clinicians' diagnoses reveals first, that the diagnosis *inappropriate vocational choice* is the most frequently-made diagnosis. Of the 35 men and six women who were diagnosed in this way, 33 men and five women did not show a primary pattern of interest on the Strong Blank in the group of occupations which included their present occupations. Reading of the case notes indicated that when the interest test data were out of line with the present vocation, the clinician almost invariably recorded the diagnosis as *inappro-*

priate vocational choice, and was unable to find any other diagnostic description more appropriate to the facts.

What diagnoses are made when the interest pattern *agrees* with the present employment? In 24 cases (14 men, 10 women) the interest test showed a primary pattern which coincided with the present job of the individual. Fourteen of these (6 men, 8 women) expressed no specific dissatisfaction. Of the 24 clients in this group, half were diagnosed as having primary personality disorders. Of the remaining 12, the diagnoses were about evenly distributed among the other categories. The following hypothesis may be formulated from these data: a person may have the vocational interests of people successfully engaged in his present occupation, but deep-seated personality disorders may otherwise interfere with his occupational adjustment.

Table 3 represents the types of treatment commonly employed by Testing Bureau clinicians. It is beyond the scope of this paper to deal with the evaluation of the various kinds of treatment. The treatment techniques most frequently used were: *placement advice* and *recommended additional training.* These were used primarily where the diagnosis was *inappropriate vocational choice.* As indicated before, these diagnoses (and therefore the treatments) were based on data from the *Strong Vocational Interest Blank. Psychotherapy* and *referral to psychiatrist* were employed in most cases which were diagnosed as primary personality difficulties.

It seems quite clear that these data allow us to test only a limited hypothesis. Actually, we are using as reference points Strong's original criterion groups. The conclusions, therefore, can only be stated tentatively until more extensive samples are utilized. If we had selected 100 adults at random among individuals who had not come to the Testing Bureau, how many would have shown primary patterns of interest which coincided with their present occupation? What proportion would have been dissatisfied with their work? What

proportion would have adjusted to such dissatisfaction without the help of an outside agency?

Partial answers to these questions are implied in a recent monograph by Darley (3). He says that individuals who continue in occupations which are at variance with their interest pattern may:

"(1) Develop socially acceptable and compensatory hobbies;
(2) Develop personality conflicts at home or on the job, but still keep on the job;
(3) Re-define the specific job duties more in line with the activities of the primary interest type . . . ;
(4) Establish a sufficiently poor work record to be only marginally employable (without promotion) or to be separated from the job . . . "

It is not improbable that this sampling is, in the main, composed of individuals who, while they may react in the alternative ways indicated by Darley, also seek the help of an available outside agency in finding an adjustment when they experience dissatisfaction.

To answer the questions raised in this discussion, crucial experiments must be carried out. Until such research is prosecuted, conclusions from these data must be made with caution. The data seem to justify this conclusion: occupational dissatisfaction is associated with a lack of primary interest in the current occupation. What explanations may be offered? Two alternatives immediately suggest themselves:

(1) A person's interests are temporally stable; they are relatively crystallized *prior* to entry into the occupational world; when the occupational activities and the interests are at variance, dissatisfaction results. The dissatisfaction is a consequent or a resultant of a fixed personality interacting within an occupational milieu.
(2) A person's interests are temporally not stable; they are flexible and subject to change *subsequent* to entry into the occupational world; they may change as a result of lack of success, environmental factors, or more fundamental personality traits in interaction. The dissatisfaction is antecedent to, or coincident with, changes from a primary pattern of interests to no

primary pattern of interests in the present occupational group.

To know which of these alternative explanations is correct is important for clinicians who are approached for assistance by vocationally-dissatisfied clients, and also by clients who are about to make a vocational choice. If interests are fixed by the time an individual is ready to seek employment, and if dissatisfaction will result if the client enters an occupation outside his interest type, then the clinician will advise him to seek employment in certain restricted areas. If, on the other hand, measured interests and satisfactions are the product of successful achievement, then the clinician will advise clients to seek employment where the greatest possibilities for success are to be found in terms of the clients' abilities and also employment opportunities. Extensive longitudinal studies will determine which of these alternatives, if either, is correct. According to Darley's review of the literature, the first alternative seems more in line with available evidence (3).

A word is in order relevant to the psychological processes which are represented by the *Strong Vocational Interest Blank*. Carter, in using this instrument, suggests that patterns of interest "become closely identified with the self." Further, "the pattern of interests is in the nature of a set of values . . ." (2). In this connection Sarbin and Berdie have demonstrated that certain relations exist between values as measured by the Allport-Vernon Scale and interests as measured by the Strong Blank (5). It is postulated that the summation of the 400 preferences on the Strong Blank reveals—at least in part—a cross-section of what the individual would like to be; in short, a person's ideal conception of the self. The Freudian expression, ego-ideal, carries approximately the same meaning.

Expressed occupational dissatisfaction, then, may be a resultant of the conflict between the ego-ideal and the occupational milieu or reality in which the individual applies this conception of the self. When the reality-situation is such that the individual's idealistic self-conception is tested and verified,

no conflict or dissatisfaction ensues. When reality prevents the testing and verification of one's ego-ideal, we find expressions of occupational dissatisfaction.

This interpretation throws no further light on the previously-posed problem as to which of the two alternative explanations is the appropriate one. The problem is merely restated in this form: is one's conception of the self (ego-ideal) a stable phenomenon or is it a variable one? Does it change with each variation in reality, with success and failure experiences? Further experimental work will illuminate some of these dark corners.

Summary

Adults who complain of occupational dissatisfaction show, in general, measured interest patterns which are not congruent with their present or modal occupations. *If* vocational interests are stable temporally, and *if* they have the dynamic character usually attributed to them, we may expect a high incidence of occupational maladjustment when individuals enter occupations for which they do not have the appropriate interests at the time of entry.

REFERENCES

1. Beane, B., Carroll, J., and Habbe, S. "The Beane Poll of Favored Psychological Tests", *Journal of Applied Psychology,* XXIV, (1940), 347-352.

2. Carter, H. D. "The Development of Vocational Attitudes", *Journal of Consulting Psychology,* IV, (1940), 185-191.

3. Darley, John G. *Clinical Aspects and Interpretation of the Strong Vocational Interest Blank.* New York: The Psychological Corporation, 1941, 71 pp.

4. Hoppock, Robert. *Job Satisfaction.* New York: Harper and Brothers, 1935.

5. Sarbin, Theodore R. and Berdie, Ralph. "The Relation of Measured Interests to the Allport-Vernon Study of Values", *Journal of Applied Psychology*, XXIV, (1940).

6. Strong, Edward K., Jr. *Manual for Vocational Interest Blank for Men*. Stanford University Press, 1940.

MEASUREMENT ASPECTS OF THE NATIONAL CLERICAL ABILITY TESTING PROGRAM

WILLIAM J. E. CRISSY
Cooperative Test Service

and

M. J. WANTMAN
University of Rochester

THE PURPOSE of the present article is threefold: to discuss the measurement procedures employed by the committee responsible for the National Clerical Ability Testing Program; to cite some of the measurement problems that confront the committee; and to suggest possible improvements in the procedures and possible solutions to the problems raised. While the organization, sponsorship, and administration of the program have been described in detail elsewhere[1] and are outside the scope of this article, it is necessary to make at least a summary statement concerning them in order to orient the reader to the subsequent discussion.

The National Clerical Ability Testing Program is sponsored by the National Office Management Association and the National Council for Business Education. Its purpose is to appraise the fitness of high school, business school, and college graduates for beginning office positions in the fields of stenography, typing, machine transcription, bookkeeping, calculating

[1] *National Clerical Ability Tests.* Bulletin No. 1, November 1939. Joint Committee on Tests. (Cambridge, Mass.: Harvard University.)

National Clerical Ability Tests. Report on 1940 Testing Program. (*ibid.* 1940.)

W. J. E. Crissy, "The Testing Program of the Joint Committee of the National Office Management Association and the Business Educational Council." *Conference of State Testing Leaders* (Proceedings of) October 28, 1939. (Washington, D. C.: American Council on Education).

machine operation, and filing. To assist in such an appraisal, the *National Clerical Ability Tests* are administered annually in centers throughout the country. In addition to tests of skill in the fields referred to, a basic battery of tests is used to measure the prospective employee's competence in English, arithmetic, business information, and general information. Certificates are awarded which are based upon the examinee's proficiency demonstrated on the special skills tests and also upon his general background as measured by the basic battery. Most examinees are candidates for just one certificate and take only one of the skills tests. However, as many as three certificates may be sought by each examinee. Certificates are awarded in each field in which the candidate reaches the required proficiency.

The procedures and problems involved in this program will be discussed under the following heads:

(1) The Separate Tests—a description of their form, content, etc.

(2) Statistical Methodology—differential weighting of each test for each group of candidates in a given skills field.

(3) Certification Procedure.

The Separate Tests

All tests in the basic battery in 1941 are of the objective type and are scored by machine. The tests in English, arithmetic, and business information are printed in a single booklet called the *Fundamentals Test*. This test requires 90 minutes of testing time. The areas of English sampled include spelling, word usage, and the use of the apostrophe in possessives and contractions. An improvement in the English Test would be the inclusion of a section measuring the examinee's knowledge of punctuation (rather than only one aspect of it) and a section testing vocabulary. The Arithmetic Test comprises problems involving the four basic arithmetic operations, and applied problems, such as the handling of discounts and the computation of interest. All items in this test are in five-choice

form. The choices include four plausible answers and a fifth alternative, *"None of the above."* The examinee must actually compute the answer to each problem since in certain items all four of the plausible answers are incorrect. The Business Information Test samples the applicant's knowledge of office procedures, postal regulations, technical business terms, and their applications. The General Information Test measures the examinee's knowledge in such areas as world affairs, sports, etiquette, geography, and history. It requires 50 minutes of testing time. This test has a wide range of discriminability and is positively correlated with every test in the battery, yet the correlations indicate some independence of measurement. It has been suggested that this test be replaced by a test of general intelligence, but various personnel officers have reported favorably on its inclusion in the battery. There is some evidence from the use of the test in employment offices to indicate that it has some value in predicting successful adjustment on the job.

The tests in the skills fields are miniature tests, that is, they present in miniature typical business situations in each of the areas included. They are long enough to measure the speed and accuracy with which particular tasks can be done over a significant period of time.

The *Stenography Test* provides for 48 minutes of dictation and 120 minutes of transcription. Fifteen items are dictated including letters and memoranda to be edited. Examinees are furnished printed copies of the letters to which the dictated letters are replies. A relatively even speed of dictation is maintained; the rate of dictation is 90 words per minute. Unusual spelling or punctuation is explained, and requests to repeat sections of the dictation are heeded if made within a specified time. The administration procedures are a departure from the usual school test in stenography but they are in accordance with usual office conditions. In the transcription similar steps have been taken to approximate business practice: erasures are permitted, change of wording is

not penalized when the sense of the item is kept, small point deductions are made for correctible errors, e.g., transposing letters, while severe penalties are imposed for uncorrectible errors such as omissions that would require interlineation. In computing the total score a bonus, proportional to the number of minutes remaining, is given to all candidates who hand in their transcriptions before "time" is called. The chief problem in connection with this test is concerned with scoring. It seems logical that, in terms of office practice, speed does not become advantageous until some acceptable level of accuracy is reached. Under the present plan of scoring, no such level of accuracy has been specified and hence some examinees obtain high scores due to their typing speed while their accuracy is at a level of doubtful acceptance in beginning office work.

The *Typing Test* permits a maximum of 120 minutes of working time. Form letters, reply cards, etc., are furnished the examinee, and he is given several typing jobs which involve the use of the materials furnished. This test approximates office conditions, as does the *Stenography Test,* and it provides a more comprehensive measure of typing ability than can be obtained from the usual kind of test in this field. However, the scoring problem is the same here as in the *Stenography Test.* No provision is made for a minimum accuracy score below which a bonus for time saved may not be added when the total score is computed.

The *Machine Transcription Test* involves the transcription of seven items from either an Ediphone or Dictaphone cylinder. A maximum of 60 minutes is allowed for the transcription. Scoring procedures are in line with business practice except, again, for the inadequate method of handling the speed aspect of the score.

The *Bookkeeping Test* requires the examinee to work to completion specified operations on excerpts from a set of books. The format of this test has been approved and recommended by experts in bookkeeping and accounting procedures. There is evidence to indicate that it measures bookkeeping

more adequately than do tests involving indirect evidences of ability in this field.

The *Machine Calculation Test* measures the examinee's ability to carry out the four basic arithmetic operations on a key-driven calculating machine. The addition problems cover columnar summing and cross-footing. The multiplication section ranges from simple multiplication to multiplying to obtain a sum of the products. The subtraction problems extend from the very simplest single subtractions through problems requiring first alternate columnar summing and then subtractions to obtain balances. The division section includes questions in direct division and also in obtaining reciprocals to be used as multipliers. The entire test requires 120 minutes of working time.

The *Filing Test* samples the examinee's ability to file various materials furnished and also his knowledge of acceptable filing procedures in the solution of problems that frequently occur in office practice. Alternative sections are included in the last part of the test to cover different filing systems. The test requires 120 minutes of working time.

Statistical Methodology

In order to obtain an over-all appraisal of the examinee's ability, scores on the basic battery and the skills test are combined into a "best-weighted" composite. Since the competencies measured by the various tests in the basic battery are of different importance in each of the six skills fields, there exist six different weighting applications, one for each group of examinees taking each of the skills tests.

Four components make up the weight accorded to each test (both basic battery tests and skills test) within each of the six fields:

(1) The variability or dispersion of the scores made by the group of examinees.
(2) A function of the test's reliability for the group; the quasi-regression weight of the test.
(3) The estimated importance of the test.
(4) The independence or uniqueness of the test.

The method of combining these components may be symbolized thus:

$$W_{ij} = \beta'_{ij} \left(\frac{I_{ij} + U_{ij}}{\sigma_{ij}} \right)$$

where

W_{ij} = weight accorded test i within a particular skills battery j (a skills battery includes a skills test and the basic battery);

β_{ij} = quasi-regression weight of the test i in battery j;

I_{ij} = estimated *importance* of the competency measured by test i in the battery j;

U_{ij} = uniqueness of test i in battery j;

σ_{ij} = standard deviation of scores on the test i of persons selecting the skills test designated for battery j.

Standard deviations are computed within skills groups for each test and are used in the weighting procedures as indicated in the formula presented above. This furnishes the first component of the weight for each test in each battery.

No criteria are now available[2] against which to obtain regression weights on the tests. However, to account for the second component, quasi-regression weights are computed using Kelley's formula[3]: $\beta' = \frac{\sqrt{r_{11}}}{1 - r_{11}}$, where r_{11} is the reliability coefficient of the test.

The third component of each test weight is determined by having a committee of experts in each skills field independently judge the importance of the competency measured by each of the basic tests (English, arithmetic, business information, and general information) relative to a basic weight of 10 accorded the skills test in that field. For example, in stenography, the *importance* weights are:

English 3
Arithmetic 1

[2]Mr. Robert Blanton has an extensive validation study in progress involving 1939 and 1940 examinees.

[3]For the derivation and rationale of this formula see: T. L. Kelley, *Interpretation of Educational Measurements* (Yonkers: World Book Company, 1927), pp. 212-213.

Business Information 2
General Information 3
Stenography 10

Uniqueness, the fourth component, is measured by using the median alienation coefficient for each test within a particular battery. This involves obtaining six intercorrelation matrices, each matrix including a particular skills test and the basic tests (5 x 5 matrix). The median correlation coefficient in each row is then used to obtain the alienation coefficient indicated above.

In order to "equalize" weights for *importance* and *uniqueness* so that the sums of the two sets of weights will be equal before the several components of the weight, W_{ij}, are combined, the alienation coefficients (uniqueness components) are each multiplied by a constant equal to the columnar sum of the five *importance* weights divided by the columnar sum of the five alienation coefficients. This is done in the case of each of the six skills batteries.

To illustrate the weighting procedure, the computational steps in the case of the 1941 Stenography battery are indicated in Table 1.

TABLE 1
STENOGRAPHY

A	B	C	C^1	D	E	F	G	H	I	J
Stenography	46.55		.90	9.487	.98	10	3.995	132.771	2.852	3.08
Bus. Inf.	11.48	.66	.59	1.873	.90	2	3.669	10.618	.925	1.00
English	7.66	.84	.70	2.789	.93	3	3.792	18.943	2.473	2.67
Bus. Arith.	4.29	.75	.73	3.164	.93	1	3.792	15.162	3.534	3.82
Gen. Inf.	22.31	.84	.82	5.031	.92	3	3.751	33.964	1.522	1.65
					$\Sigma E=4.66$	$\Sigma F=19$				

Column *Data Presented*

A Tests included in Stenography battery.

B Standard deviation for each test within the Stenography group (σ_{ij}).

C Reliability coefficients computed by "split-half" method based upon sample.

C[1] Adjusted reliability coefficients (corrected to range of Stenography group) by formula:

$$r_{11} = 1 - \frac{\Sigma^2}{\sigma^2}(1 - R_{11})$$

D Quasi-regression weights (β'_{ij}).

E Uniqueness weights (median k's).

F Judges' weights of importance (I_{ij}).

G Entries in column E adjusted relative to entries in column F by the formula:

$$G = \frac{\Sigma F}{\Sigma E} E = (U_{ii})$$

H $D(F + G) = \beta'_{ij}\ (I_{ij} + U_{ij})$

I $$\frac{H}{B} = \frac{\beta'_{ii}\ (I_{ii} + U_{ii})}{\sigma_{ij}} = W_{ii}$$

J Entries in I, each divided by smallest entry in I (Column J contains the weights which are finally used. This makes subsequent computation easier by making the smallest W_{ij} equal to unity.)

When the weight for each test has been obtained by the foregoing procedures, a composite score for each examinee is obtained by multiplying each score by the appropriate weight and summing his weighted scores.

Certification Procedure

The certification of an examinee in a particular skills field depends upon two factors (1) having a skills test score equal to or in excess of the critical or "passing" score set for that particular skills test; and, in addition, (2) having a composite score equal to or in excess of the critical or "passing" score set for that particular composite.

The critical score for each skills test is established by a committee of experts in that field (usually the same committee that estimates the *importance* weights). The criterion used by each committee is "minimum acceptable performance in a beginning office job." The procedure used is to have each committee member inspect and judge as acceptable or unacceptable a sample paper from each two per cent segment of the

distribution beginning at the twentieth percentile point and extending to the eightieth percentile point. After independent judgments are completed, the combined judgment of the committee is used to determine the critical score on the particular skills test.

To determine the critical composite score, regression technique is employed; the desired critical composite score is predicted from the previously determined critical skills score through regression of composite on skill.

Obviously the correlation between each skills test and the corresponding composite is high, since the skills test is the chief component of the composite. An improvement in this procedure would be to exclude the skills test from the composite and thus use the composite as a "background index." Then if the critical composite score were obtained by prediction from the critical skills test score by means of a regression equation involving composite and skill, the only hypothesis involved would be that the minimum "background" score should be about equivalent to the minimum skills score.

Summary

In this paper have been discussed the measurement procedures and problems connected with the National Clerical Ability Testing Program. The treatment of problems has been limited to those which are peculiar to this particular program. The procedures, however, have been covered in detail because they have general application to other types of testing projects.

The weighting and certification procedures described in this paper should obviously be completely revised as soon as "outside criteria" are available against which to weight the tests. Probably the best procedure to use when these criteria are available is canonical correlation technique modified to include *importance* weightings of both the separate criteria and the separate tests.

So long as such outside criteria are not available, the pro-

cedures used at present should be modified either in accordance with the suggestions made in this paper or in some other manner if the program is to render increased service to prospective clerical employees and to the employers of such persons.

INTROVERSION-EXTROVERSION AS A FACTOR IN TEACHER-TRAINING

CATHARINE EVANS
Indiana University
and
C. GILBERT WRENN
University of Minnesota

Introduction

ONE OF THE SERIOUS problems facing teacher-training institutions today is the selection of student personnel. Many teachers are unsatisfactory either because of inadequate training or unfortunate personality characteristics. Teacher-training institutions must select students who can benefit most from the improved training programs now provided. This study is intended to throw some light on the relationship of personality traits to student success in teacher-training programs. More specifically, the purpose of this study has been to determine the relationship of Introversion-Extroversion[1] to the scholastic achievement and student teaching success of education students. An *I-E Inventory* was administered to 396 seniors in the College of Education at the University of Minnesota. This inventory will be described briefly in order that the results of the study may be understood clearly. A more complete discussion of the construction of the inventory is available in a recent article in the *Journal of Psychology* (1).

This inventory was constructed to measure three types of

[1]Throughout the remainder of the article, Introversion-Extroversion will be designated as I-E.

I-E, Thinking, Social, and Emotional, which were isolated by Guilford (2) in his factor analysis of I-E. Original items were developed and stated in the form of questions concerning the behavior and reactions of the student. The questions were formulated in such a manner that the student could indicate how frequently he or she behaved in that way. Typical questions were, "Do you question statements and ideas expressed by your professors?" "Do you enjoy eating meals alone?" and "Do you avoid exaggeration in your statements?"

The construction and choice of items for the three tests in the inventory were guided by the following definitions which contrast the extremes for each type of I-E:

The thinking introvert likes reflective thought, particularly that of a more abstract nature. His thinking tends to be less dominated by objective conditions and generally accepted ideas than thinking of the extrovert. The thinking extrovert, however, shows a liking for overt action, and his ideas tend to be ideas of overt action.

The social introvert withdraws from social contacts and responsibilities and displays little interest in people. In contrast, the social extrovert seeks social contacts and depends upon them for his satisfaction.

The emotional introvert tends to repress and inhibit outward expression of his emotions and feelings. On the other hand, the emotional extrovert readily expresses his emotions and feelings outwardly. He shows a greater tendency to make the expected response to simple, direct emotional appeals than the introvert.

In constructing this inventory an effort was made to develop relatively independent measures for these three types of I-E by a technique of item analysis. The intercorrelation coefficients for the 396 College of Education seniors were: Thinking and Social I-E tests, —.25; Thinking and Emotional I-E test, +.17, and Social and Emotional I-E tests, + .23. These low coefficients indicate that the three tests are relatively independent. The inventory also seems sufficiently reliable for individual prediction since each of the tests has a reliability coefficient with groups of education students of .88 or above,

for either the retest or split-half technique or for both techniques.

Indirect evidence concerning the validity of each test is available, in terms of the ability of the test to differentiate known groups of college students which on an *a priori* basis seemed to be extreme in the type of I-E involved. For example, the Thinking I-E test significantly differentiated major groups in the College of Education; the majors in physical education, home economics, commercial education, and child welfare were extreme in the direction of Thinking Extroversion, while the majors in English, art, mathematics, social studies, and languages were extreme in the direction of Thinking Introversion. The Social I-E test significantly differentiated groups of students varying in the degree of participation in campus activities; the members of academic sororities and fraternities and the students active in campus organizations were found to be more socially extroverted than the non-affiliates and non-participants. Likewise, the Emotional I-E test significantly differentiated sex and age groups; women were more emotionally extroverted than men, and the younger student groups were more emotionally extroverted than older groups. Each test did differentiate known groups of students which logically were expected to be extreme in that type of I-E.

Scholastic Achievement of Student Groups Varying in Thinking I-E

The relationships of the scores on the three tests in the *I-E Inventory* to measures of scholastic and student teaching success have been explored in this study of College of Education seniors.

The relationship of Thinking I-E to scholastic achievement honor point ratios was explored for the seniors in the College of Education who had taken a scholastic aptitude test, the *Miller Analogies Test,* during the junior or senior year. The *Miller Analogies Test*—Form G, consists of 100 analogies which research with college students indicated were discriminating items. Data reported by Dugan (3) indicate that

this test is highly reliable and valid as a measure of scholastic aptitude. The correlation between the scores on the Thinking I-E and the *Miller Analogies Tests* varied from .15 to .26 with groups of 112 to 260 education students, indicating only a small positive relationship.

Significant results were obtained in the study of the scholastic achievement of 148 native students, i.e., students who had taken all of their college work at the University of Minnesota. These students were divided into four groups according to the varying size of their scores on the Thinking I-E test, and the mean honor point ratios for each group were computed for the total course work, for major courses, and for education courses. There was in general a progressive increase in the mean major and education honor point ratios with an increase in introversion in this group. Those students with Thinking I-E scores in the quarter extreme in the direction of introversion had significantly higher mean honor point ratios than those with scores in the quarter extreme in the direction of extroversion (see Table 1).

TABLE 1

MEAN HONOR POINT RATIOS IN TOTAL, MAJOR, AND EDUCATION COURSES FOR FOUR GROUPS OF NATIVE STUDENTS VARYING IN DEGREE OF THINKING I-E

	Mean Honor Point Ratio		
Thinking I-E Groups	Total	Major	Education
Upper Quarter (Extroversion)	1.3711	1.6743	1.5557
Second Quarter	1.5211	1.7878	1.8050
Third Quarter	1.6776	1.9051	1.8849
Lowest Quarter (Introversion)	1.6138	1.9292	1.8926

TABLE 2

DIFFERENCE AND THE SIGNIFICANCE OF THE DIFFERENCE IN MEAN HONOR POINT RATIOS FOR THE EXTREME THINKING I-E QUARTILE GROUPS OF NATIVE STUDENTS

Variable	Difference in means	t	Probability of t
Total Honor Point Ratio	.2427	2.9241	.01
Major Honor Point Ratio	.2549	2.7117	.01
Education Honor Point Ratio	.3369	2.9553	.01

The analysis of variance technique (4) was applied to the total, major, and education honor point ratios for the four Thinking I-E groups in order to test the significance of the difference in means. The variance among the mean honor point ratios of the Thinking I-E groups was significantly larger for each of the three analyses of variance than the variance within the groups. The ratio of the variances for each type of honor point ratio satisfied at least the five per cent level of significance. The results of the analysis of variance for each type of honor point ratio refuted the hypothesis that there was no difference in the scholastic achievement of the four Thinking I-E groups. The groups were heterogeneous in scholastic achievement (see Table 2).

The analysis of variance technique with the two criteria of classification, Thinking I-E and *Miller Analogies* scores, was also employed with the total honor point ratios in order to determine whether or not the variance among the mean honor point ratios of the four Thinking I-E groups would be significant when the groups were subdivided according to Analogies ability.

The variance between the Thinking I-E groups was not significantly greater than the variance within the Thinking-Analogies subclasses. When the variance within the Analogies quartile groups was considered, there was insufficient evidence to determine any difference in the scholastic achievement of the four Thinking I-E quartile groups.

From the analysis of variance data the mean honor point ratios of the following four groups of native students varying both in degree of Thinking I-E Analogies ability were compared:

(1) Students below the median in the direction of Thinking Introversion and above the median in the ability measured by the Analogies Test.

(2) Students above the median in the direction of Thinking Extroversion and above the median in Analogies ability.

(3) Students below the median in the direction of Thinking Introversion and below the median in Analogies ability.
(4) Students above the median in the direction of Thinking Extroversion and below the median in Analogies ability.

A steady decrease in mean total honor point ratio from the first to the fourth groups can be noted in Table 3.

TABLE 3

MEAN TOTAL HONOR POINT RATIOS FOR FOUR GROUPS OF NATIVE STUDENTS VARYING IN THE DEGREE OF THINKING I-E AND OF THE ABILITY MEASURED BY THE MILLER ANALOGIES TEST

Number of Group	Type of Groups	Number	Mean Total Honor Point Ratio
1	*Below* the median in the direction of Thinking Introversion and *above* the median in Analogies ability.....	46	1.7565
2	*Above* the median in the direction of Thinking Extroversion and *above* the median in Analogies ability.....	28	1.5218
3	*Below* the median in the direction of Thinking Introversion and *below* the median in Analogies ability.....	28	1.4636
4	*Above* the median in the direction of Thinking Extroversion and *below* the median in Analogies ability.....	46	1.4000

The first group had a significantly higher mean honor point ratio than the second group ($t = 2.24$). However, this second group did not have a significantly higher mean honor point ratio than the third ($t = .58$). These data seem to indicate that high Analogies ability combined with a tendency toward Thinking Introversion is more ideal from the standpoint of scholastic achievement than either high Analogies ability combined with a tendency toward Thinking Extroversion or low Analogies ability combined with a tendency toward Thinking Introversion.

The question of the relation of Thinking Introversion and high Analogies ability to scholastic achievement was attacked

from another angle. The 24 native students who had honor point ratios above 2.00, and the 26 who had honor point ratios below 1.50 were compared in mean scores on the Thinking and Analogies Tests (Table 4). The students with the higher honor point ratios had significantly higher Thinking Introversion scores and Analogies scores than those students with lower honor point ratios. The values of "*t*" were 2.51 and 3.58, respectively, and they satisfied at least the five per cent level of significance.

TABLE 4

MEAN THINKING AND ANALOGIES SCORES OF NATIVE STUDENTS WITH HONOR POINT RATIOS ABOVE 2.00 AND OF NATIVE STUDENTS WITH HONOR POINT RATIOS BELOW 1.50

Type of Group	No.	Mean Thinking I-E Score*	Mean Analogies Score
Natives with Honor Point Ratios above 2.00	24	6.8333	60.0833
Natives with Honor Point Ratios below 1.50	26	23.4615	45.6538

*The smaller the score, the greater the tendency toward introversion.

The evidence for the relationship of Thinking Introversion to scholastic achievement for native students seems weakened by the non-significant results obtained in the analysis of variance by the double criteria of classification. However, the study of group differences points to the desirability of the combination of a tendency to Thinking Introversion with high Analogies ability. Indeed, the results of the analysis of variance by the double criteria of Thinking I-E and Analogies ability can be interpreted as strengthening the evidence for the greater desirability of the combination of Thinking Introversion with high Analogies ability in contrast to the desirability of a tendency to Thinking Introversion alone regardless of Analogies ability.

When transfer students, i.e., those students transferring to the University of Minnesota with advanced standing from

other institutions, were studied, the results were not significant. Although the mean honor point ratios of the transfer students in the quartile group extreme in the direction of Thinking Introversion were larger than the means of those in the quarter extreme in the direction of extroversion, the differences were not significant. Likewise, the results of the analysis of variance were not significant. However, significant differences were found between samples of native and transfer students in the distribution by major fields, in the mean scores on the Analogies test and in the means of the major and education honor point ratios. The explanation of the contrasting results obtained with native and transfer students must lie in these differences. It may be safely concluded however, that there is a relationship between Thinking Introversion and scholastic success for the native student in the College of Education.

The Student Teaching Success of Groups of Students Varying in I-E

The relationships of scores on the three I-E tests to student teaching success at the University of Minnesota were also explored. Two measures of student teaching success were employed in this study. In the first place, the rank order lists of the student teachers in 16 major fields were obtained. These rank order lists were made out by the combined group of critic teachers for a major field. Second, the marks in student teaching for the three quarters were computed as an honor point ratio for 242 seniors.

The coefficients of correlation between the student teaching ranks and the ranks of the scores on the three I-E tests were computed for the majors in eight of the 16 teaching fields. These were the majors which numbered 14 or more cases. The Spearman rank difference formula was employed in the calculation of these rank correlation coefficients.

These coefficients as given in Table 5 varied from .00 to .43. They were based on such small samplings that it was improbable that any of the coefficients were significant. All of the coefficients of correlation between Social I-E and student

TABLE 5

RANK COEFFICIENTS OF CORRELATION BETWEEN STUDENT TEACHING RANKS AND THE THREE I-E TESTS FOR EIGHT MAJOR FIELDS

Major Field	*N*	Thinking I-E Test	Social I-E Test	Emotional I-E Test
1. English	50	—.23	+.19	+.24
2. Social Science	49	—.06	+.17	+.23
3. Child Welfare	38	—.20	+.12	—.03
4. Science	19	—.04	+.38	—.38
5. Commercial	19	—.18	+.43	—.12
6. Art	17	+.37	+.19	+.16
7. Girl's Physical Education	15	—.37	+.35	—.41
8. Mathematics	14	.00	+.25	+.18

teaching rank, however, were positive. This consistent positive relationship of Social Extroversion and student teaching rank for the eight major fields does seem to indicate the type of relationship existing between these two factors. Six of the eight coefficients between student teaching ranks and the scores on the *Thinking I-E Test* were negative. Thus a tendency was indicated for Thinking Introversion to be related to student teaching success. The relationship of Emotional I-E to student teaching rank was not consistent in the eight major fields. Four coefficients were positive, and four were negative.

The relationship of student teaching success to the I-E scores was also determined for 55 students whose student teaching ranks were in the upper and lower quarters on the rank order lists of the same eight major fields. This analysis indicated that the more successful student teachers tended toward more thinking introversion, social extroversion, and emotional extroversion in central tendency than the less successful student teachers. However, only one of these differences in mean I-E scores was significant. The students in the upper fourth on student teaching rank order lists were significantly more socially extroverted than the students in the lower fourth; the value of "*t*," 2.58, satisfied the five per cent level of significance. There were, on the other hand, more than five chances out of one hundred that the differences in the

mean Emotional and Thinking scores could have occurred from chance errors of sampling.

Since there was a significant difference in the mean scores on the Social I-E test for the groups in the upper and lower fourths on the student teaching rank order lists, the analysis of variance technique was employed to study these differences. The mean scores on the Social I-E test of the four groups formed on the basis of ranks in student teaching are given in Table 6. There was a progressive increase in mean scores on

TABLE 6

MEAN SCORES ON THE SOCIAL I-E TEST OF THE FOUR GROUPS FORMED ON THE BASIS OF RANK IN STUDENT TEACHING

Groups Varying in Student Teaching Rank	No.	Mean Social Score*
Upper Quarter	55	15.9273
Second Quarter	57	15.2632
Third Quarter	54	7.7963
Lower Quarter	55	4.0727

*The larger the score, the greater the degree of extroversion.

the Social I-E test for the four groups with the increase in student teaching rank. In other words, the tendency to Social Extroversion increased as student teaching rank became higher. The simple analysis of variance was employed to test the null hypothesis that there was no difference in the four groups from which the mean Social I-E scores were obtained. The variance among the four groups varying in practice teaching ranks was significantly greater than the variance within the groups, indicating that the four groups were not homogenous in Social I-E. These data indicated that Social Extroversion was related to student teaching success as measured by the ranks of critic teachers. The more successful the student teacher, the greater was the tendency to Social Extroversion.

The marks in student teaching for three quarters of work were also employed as a criterion for the choice of two groups of students of extreme degrees of success in teaching. A

majority of the seniors had student teaching honor point ratios between 2.00 and 2.50. There were 68 students with ratios above 2.50 and 67 students with ratios below 2.00.

The mean scores on the I-E tests of these two groups were compared. The same differences were found for these two groups as for the two groups chosen by the criterion of student teaching ranks with the exception of the scores on the Thinking I-E test. The mean scores indicated that the more successful student teachers as judged by their marks were more extroverted in Thinking, Social, and Emotional I-E than the less successful teachers. However, the values of "*t*" were so small that none of the three differences in means was significant. The rank order lists seemed to yield a more differentiating measure of student teaching success than marks in student teaching. In fact, students with a B average (2.00 ratio) in student teaching varied in the rank given by critic teachers from the lowest to the upper quarter.

The results of the use of the two criteria for teaching success indicate that the more successful student teacher is on the average more extroverted, socially and emotionally, than the less successful student teacher. No conclusion can be made in relation to the Thinking I-E test because of the conflicting results.

Summary

The results of this study indicate that for "native" seniors in the College of Education, Thinking Introversion is related to high scholastic achievement and that Social and Emotional Extroversion are related to student teaching success. These results provide evidence to substantiate a common "hunch" that good teachers are not only good students but also must possess certain social and emotional characteristics. The results indicate that a combination of high mental ability and Thinking Introversion is desirable for scholastic success. In addition, Social Extroversion is also necessary for high rank in student teaching. The extent to which the I-E Inventory[2]

[2]This inventory will be published soon by Science Research Associates.

can be used to predict scholastic success in the College of Education or student teaching success has not yet been determined. The I-E Inventory is being used, however, in a current study in this same college that will follow a class of juniors through a four-year period. By studying their success in student teaching and on the job, it will be possible to indicate the predictive values of the inventory and other instruments not only for the training period but also for job adjustment.

REFERENCES

1. C. Evans and T. R. McConnell. "A New Measure of Introversion-Extroversion," *Journal of Psychology*, XII (1941), 111-124.
2. J. P. Guilford and R. B. Guilford. "Personality Factors, S, E, and M, and Their Measurement." *Journal of Psychology*, II (1936), 109-127.
3. Willis E. Dugan. "A Study of the Miller Analogies Test with Graduate Students in the College of Education." Unpublished Master's Thesis. University of Minnesota, 1939.
4. George W. Snedecor. *Statistical Methods*. Ames, Iowa: Collegiate Press Inc., 1938, 387 pp.

AN INVESTIGATION OF THE POSSIBILITIES OF MEASURING PERSONALITY TRAITS WITH THE STRONG VOCATIONAL INTEREST BLANK[1]

LYLE TUSSING

Wilson Junior College

THE *Strong Vocational Interest Blank* is a widely used instrument for determining vocational interest patterns in educational and vocational guidance. This study was contemplated because it was believed that there was a possibility of weighting items on the *Interest Blank* in such a way as to obtain, with this single test, certain personality measures as well as the vocational interest scores now available. With this object in mind, the present study analyzes the relationship between responses on the *Strong Vocational Interest Blank* and scores on certain personality tests to determine how well the traits measured by these tests can be measured by the Strong Blank.

The idea of evaluating other factors than vocational interest with the *Strong Vocational Interest Blank* is not new. Strong has used his Blank to measure masculinity and femininity (20) and also interest maturity (18, 19). Young and Estabrooks (21) studied the relation of personality and interest tests to scholastic success. They found that the *Strong Vocational Interest Blank* showed evidence of being the most significant predictive measure after they had made an item analysis of several tests.

[1]This study is a portion of a thesis submitted to the Faculty of Purdue University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, June, 1941.

The purpose of the present study was to construct scoring keys for the *Strong Vocational Interest Blank* to measure certain personality traits, such keys being: (1) suitable for use in group testing and available as another key for the *Strong Vocational Interest Blank,* (2) so constructed that scoring is both rapid and free from the personal error. Validation of the scales was obtained by correlating the scores resulting from the new Strong scoring keys with scores for the corresponding Allport-Vernon, Bell, Bernreuter, and the American Council on Education tests.

The matter of falsification of responses to test items is a problem that is worthy of consideration. With most personality tests, it is a very easy matter for the person being tested to falsify his responses if he wishes. (16) It has also been shown that the *Strong Vocational Interest Blank* can be changed to falsify interest in a vocation (13). However, in most cases where the individual is interested in obtaining guidance, he will not deliberately falsify his responses. Also it seems that because of the number and brevity of the items in the *Strong Vocational Interest Blank,* falsification of responses by the subject possibly would be more difficult than in the conventional personality inventory of the direct question type.

It was not deemed necessary in the course of this investigation to examine the nature of personality traits, to define them, nor to investigate all the possible traits. In this study, the validity and reliability of the tests used was accepted, and no attempt was made to prove or disprove whether the tests selected were valid measures of the traits which they purported to measure. (1-10 and 15)

The exact number and the names of existent personality traits have not been agreed upon. Allport (4) states that "The Eugenics Record Office has issued a *Trait Book* containing a list of approximately 3,000 characteristics that might conceivably be hereditary according to the principle of unitary characteristics." Further he cites McDougall as listing five elements of personality; Beck, four; and Boven, three. While

it is unlikely that authorities will agree as to the units of personality and their exact number, nevertheless, measures of "sociability", "confidence in one's self", "home adjustment", "health adjustment", and "emotional adjustment", as well as intelligence, are quite widely measured factors of personality, and consequently these measures have been selected for investigation in the present study.

Procedure

In this study the problem was to determine how the items on the *Strong Vocational Interest Blank* were related to the scores made on the Allport-Vernon, Bell, and Bernreuter personality tests, and on intelligence tests, and also to find whether the Strong Blank could be used as a means of measuring the same personality traits as the above-mentioned tests if the items were weighted. In order to obtain weightings, it was necessary to determine how the groups scoring at the extremes of the measures (high, low) varied in their responses to items on the *Strong Vocational Interest Blank*.

A sample group of 300 men was used for establishing the weights. This group was one originally studied by Dr. E. Lowell Kelly (14) in an investigation of factors contributing to marital happiness.[2] Consequently, it was a group in which several selective factors were operating. For example, each man was about to be married. He was willing to cooperate in a study in which such factors as intelligent curiosity and cooperativeness play an important role. His general intelligence and educational background were perhaps higher than the average. Most of this group had attended college, and the average age of the group was 26.66 with a standard deviation of 3.47. However, even though these selective factors were operative, the group nevertheless represented a relatively heterogeneous group with respect to the variables under analysis and therefore was satisfactory for the present investigation.

[2]Data for these studies were collected with the aid of a series of grants from the Committee for Research on Problems of Sex of the National Research Council.

The 300 subjects were given a battery of tests including the *Strong Vocational Interest Blank,* the Allport-Vernon *Study of Values,* the Bernreuter *Personality Inventory,* the Bell *Adjustment Inventory,* and the *Otis S-A Test of Mental Ability.* The present study utilizes their responses to these tests as basic data.

Scores of the 300 men on the nine scales (four scales of the Bell *Adjustment Inventory;* two of the Bernreuter *Personality Inventory,* F1-C and F2-S; the *Otis S-A Test of Mental Ability;* two scales, theoretical and economic, of the Allport-Vernon *Study of Values*) were coded and punched on cards. This set of cards was then sorted on each of the nine scales to identify the subjects scoring in the highest 25 per cent and the lowest 25 per cent of the scores on each scale. In many instances, due to coding, it was impossible to select exactly 25 per cent (75 cases), although this number was desired in each of the two groups.

After these two extreme scoring groups were obtained, the next step was that of comparing the responses of these groups for each item on the Strong Blank. This was done by punching the responses of the 300 subjects to the 420 items of the 1927 form of the *Strong Vocational Interest Blank* on Powers cards. This punching consisted of recording the responses for each individual on each item as to "like", "indifferent", or "dislike" for the particular item. In some instances the subject failed to respond to an item, but, since inspection showed these omissions to be fairly equally distributed in the high and low groups, it was considered advisable not to weight such omissions.

After determining which individuals constituted the high and low groups of a particular test, the cards containing the Strong responses of the individuals composing each group were hand sorted by serial number of the subject. Each resulting "high" group of cards was then sorted by item and the number of persons answering "like", "indifferent", and "dislike" for each of the 420 items in the high group was recorded. This

same procedure was followed in the low group. The percentages were then obtained for each of the 420 items for "like", "indifferent", and "dislike" for both the high and low groups (2,620 percentages for each scale). This procedure was followed for the nine scales.

After the percentages had been computed for the high and low groups on "like", "indifferent", and "dislike", the difference in the two percentages for "like", "indifferent", and "dislike" for each item was compared and the differences in percentage were given a weighting. These weightings ranged from +4 to —4 according to the formula $W = 100 \frac{\Delta}{1-4j\Delta^2}$ reported by Strong (12, 17) to be the most satisfactory scheme he had found for assigning weights to his interest scales.

These new weights were then transferred to the 1938 Strong Blank. Since the 1938 form included only 400 of the original 420 items, some of the responses had to be omitted in making this transfer. However, the number of changes was small and probably did not appreciably affect the reliability or validity of the new scales. The weights as they now appeared on the 1938 form of the *Strong Vocational Interest Blank* constituted a new scoring key. These weights were then transferred to International Business Machines test-scoring keys.

A new sample of 103 male college freshmen and sophomores was used as a validating group. These students took the battery of tests consisting of the *Strong Vocational Interest Blank,* the Allport-Vernon *Study of Values,* the Bernreuter *Personality Inventory,* the Bell *Adjustment Inventory,* and the American Council on Education *Psychological Examination* voluntarily. Since considerable time was consumed by each individual participating, not all of the men finished all of the tests in the battery. However, as many subjects as possible were utilized in the validation of each new scoring key.

The responses of the validating group on the *Strong Vocational Interest Blank* were then scored with each of the various

new keys (i.e., the key for "economic",[3] "theoretical", "intelligence", etc.). By scoring the odd and even items separately, it was possible to compute the reliability for each new scale. The scores made on the "odds" were correlated with those made on the "evens", and the reliability for the whole test was estimated by means of the Spearman-Brown formula.

The validity was obtained by correlating the total (odd plus even) scores on the *Strong Vocational Interest Blank;* for example, the total scores on "theoretical" with the theoretical scores obtained on the Allport-Vernon *Study of Values.*

Results

Table 1 indicates the number of weighted items on the *Strong Vocational Interest Blank.* It also shows the results of the validating group for the Bernreuter (F1-C, F2-S Scales), Allport-Vernon (Theoretical, Economic Scales), Bell (Home, Health, Social and Emotional Adjustment Scales), and Intelligence as measured by the American Council on Education *Psychological Examination.*

TABLE 1

RESULTS OF THE NEW STRONG BLANK SCALES

	"F1-C"	"F2-S"	"Theor."	"Econ."	"Home Adj."	"Health Adj."	"Social Adj."	"Emot. Adj."	"Intel."
Mean	−17.4	19.5	9.3	4.2	−12.8	−5.0	48.5	−6.7	23.7
S. D.	58.2	36.1	64.3	64.9	38.2	22.2	54.6	25.3	53.4
Reliability	.86	.77	.82	.80	.85	.70	.87	.77	.80
Validity	.48	.52	.48	.56	.21	.34	.50	.07	.45
No. Wgtd. Items	347	327	336	325	319	246	339	289	343
No. Items Wgtd. 2 or more	130	70	143	118	54	24	108	34	129

[3]To distinguish scores on the new from those on the original scales, each of the newly derived scales is designated by using the same name or symbol as for the original test, but is enclosed in quotation marks.

It is interesting to note the items on the Strong Blank which compose these scales. Some of the better (more heavily weighted) items on the Strong "F1-C" Scale (The Bernreuter F1-C Scale purports to differentiate the self-conscious individual from the self-confident individual) which appeared to differentiate the *self-conscious* individual from the *self-confident* person show that the former dislikes the occupation of "aviator." He dislikes the school subjects "physical training" and "public speaking." In the field of amusements, he likes "picnics", but dislikes "rough house invitations." With reference to activities, he dislikes "adjusting a carburetor", "repairing electric wiring", "making a speech", "organizing a play", "opening a conversation with a stranger", "acting as a yell-leader", and "expressing judgments publicly regardless of criticism." In the section dealing with peculiarities of people, he dislikes "people who assume leadership", but likes "talkative people." He would rather listen to a story than tell a story. In rating his present abilities and characteristics, he does not usually start the activities of a group, work steadily, or liven the group on a dull day. He is not quite sure of himself, and he is doubtful as to his ability to accept criticism without getting sore and to distinguish between more or less important matters. He does not discuss his ideals with others, and his feelings are easily hurt. He loses his temper at times and worries considerably about mistakes.

Opposed to this, the self-confident individual likes "meeting and directing people", but he dislikes "talkative people." He is indifferent as to whether he would rather tell a story or listen to a story, and he likewise has no preference in the matter of a few intimate friends or many acquaintances. He answers in the affirmative to "am quite sure of myself", to "accept criticism without getting sore", and "discuss my ideals with others." He does not get rattled easily, and his feelings are not easily hurt. The person who is self-confident "enters into the situation and enthusiastically carries out the program", and he does not worry.

The Bernreuter F2-S Scale is a measure of sociability. Some of the better (i.e., more heavily weighted) items on the Strong "F2-S" which appeared to differentiate the *non-social* individual from the *social* individual show that the non-social individual likes to "deal with things" rather than people. He would like to be a magazine writer, and he prefers "amusements alone or with two or three others" rather than "amusement where there is a crowd." He would rather spend nights at home than away from home, and he enjoys reading a book rather than going to the movies. The non-social individual reports that he has a "few intimate friends" rather than "many acquaintances." He says he does not "win friends easily", nor does he "usually liven up the group on a dull day." He "can write a concise, well-organized report." Such an individual "practically never tells jokes" and his "feelings are easily hurt."

The social individual is one who "tells jokes well." He is also a person who is indifferent in his choice between a "great variety of work" or a "similarity of work."

The Allport-Vernon Theoretical Scale was designed to segregate individuals whose theoretical values in life are highest. On the new Strong "Theoretical" Scale, there are a number of more heavily weighted items which appear to differentiate the *theoretical* individual from the *non-theoretical* type. For example, in indicating the occupations he would prefer, the former names as "likes" the following: "astronomer", "author of a technical book", "chemist", "inventor", "laboratory technician", "marine engineer", "scientific research worker", "statistician", and "watchmaker." The theoretical person is fairly consistent in preferring such school subjects as "calculus", "geology", "philosophy", "physics", "physiology", and "zoology." The theoretical type of individual enjoys "museums" and the "solving of mechanical puzzles", also "doing research work." In the section of the Strong Blank devoted to the order of preference of activities, he indicates "like" for the development of the theory of operation of a

new machine and for the role of chairman of an educational committee. He would have liked to have been a "Luther Burbank" or a "Thomas Edison." He likes "technical responsibility (head of a department of 25 people engaged in technical, research work)" as opposed to "supervisory responsibility (head of a department of 300 people engaged in typical business operation)", and he would prefer "mental activity" to "physical activity." In rating his own abilities and characteristics, the theoretical person checks in the affirmative that he has mechanical ingenuity. In the division of pecularities of people, he dislikes "bolshevists."

A few of the items which were more heavily weighted to characterize the non-theoretical individual are that he would like to be a "sales manager" and would have liked to have been "John Wanamaker, merchant." He dislikes the subject of "chemistry."

The economic man as described by Allport (6) is "characteristically interested in what is useful. Based originally upon the satisfaction of bodily needs (self-preservation), the interest in utilities develops to embrace the practical affairs of the business world. This type is thoroughly practical and conforms well to the prevailing stereotype of the average American business man."

On the "Economic" Scale of the *Strong Vocational Interest Blank,* some of the more heavily weighted items which appeared to differentiate the *economic* man from the *non-economic* man indicate that he likes the occupation of "sales manager" and likes to "develop business systems." He would have liked being "Henry Ford, manufacturer", "J. P. Morgan, financier", or "John Wanamaker, merchant." This "typical business man" is indifferent toward the pastime of "reading a book" as compared with that of "going to the movies." He definitely dislikes the occupations of "artist" and "author of a novel", and has the same attitude toward the subject of "literature." He likewise records dislike for "absent-minded people", "socialists", and "bolshevists." He would

prefer "supervisory responsibility (head of a department of 300 people engaged in typical business operation)" to "technical responsibility."

In contrast, the non-economic man indicates that he likes such occupations as "clergyman", "college professor", "magazine writer", "poet", "school teacher", "sculptor", and "social worker." In the same vein, the non-economic person likes the subject "art" and likes "poetry." He admires "Luther Burbank, plant wizard", "Charles Dana Gibson, artist", and "Booth Tarkington, author." He likes amusements such as "observing birds", "visiting art galleries", and "museums", and listening to "symphony concerts."

In investigating the Bell *Adjustment Inventory,* the scales for "Home Adjustment", "Health Adjustment", and "Emotional Adjustment" on the Strong Blank showed relatively few weightings of two or more. This dearth of heavily weighted items and the large number of unit weight items in no apparent pattern, seems to indicate the inability of the Strong items to differentiate people scoring high from those scoring low on the Bell Home, Health, and Emotional Adjustment Scales. This fact is substantiated by the very low validity coefficient obtained.

Some of the more heavily weighted items which appear on the "social adjustment" scale that differentiate the *social* from the *non-social* type of person show that the former chose an occupation such as "consul." He liked subjects such as "dramatics", "literature", and "public speaking." The socially adjusted person indicated a liking for activities such as "interviewing clients", "opening a conversation with a stranger", "making a speech", "organizing a play", "meeting and directing people", "taking responsibility", "meeting new situations", and "entertaining others." He stated that he liked "quick tempered people" and that he "tells jokes well." He answered the following in the affirmative: "usually start activities of my group", "usually get other people to do what I want done", "usually liven up the group on a dull day", and "am quite sure of myself."

The non-social individual preferred to be a "member of a society" rather than an "officer in a society." He would rather "deal with things" than people. He would choose a "few intimate friends" rather than "many acquaintances." He would rather "listen to a story" than "tell a story." He stated that he "worries considerably about mistakes", his "feelings are easily hurt", and his "advice is practically never asked." The non-social person indicated a dislike for a "politician" and for "expressing judgments regardless of criticism."

An investigation of intelligence as measured by the American Council on Education *Psychological Examination* shows the following more heavily weighted items which appear to differentiate the person of *high intelligence* from the individual with a low intelligence rating. The former indicates a liking for the occupations of "author of a novel", "college professor", "editor", and "magazine writer." He dislikes the occupations of "life insurance salesman" and "office clerk." The person of high intelligence indicates a preference for "symphony concerts" and "poetry." He would rather read a book than go to a movie. He chooses the "Atlantic Monthly" for reading and prefers the school subjects of "algebra", "calculus", "geometry", "literature", and "philosophy." He enjoys "arguments", the "teaching of adults", and admires "independents in politics." He considers the most important factor affecting his work an "opportunity to make use of all knowledge and experience", and he says that he "can write a concise, well-organized report."

The occupation "floorwalker", the amusement "vaudeville", the subject "agriculture", and the man "John Wanamaker, merchant" are items which the individual of low intelligence rates as liked. He dislikes the occupations of "astronomer", "author of a technical book", "inventor", and the subject of "sociology." This dislike is also evidenced in the weighting of the items "Bolshevists", "writing personal letters", and "Booth Tarkington, author." The person with a low intelligence rating indicates that he would prefer to

"work for self in small business" than to "work in a large corporation with little chance of becoming president until age 55"; also, that he likes "many acquaintances" rather than a "few intimate friends."

Additional information was obtained on some of the new scales. A correlation was found between grade point averages for the first semester in school and the scores on the intelligence scale. The correlation was .42 (N = 79). The validity of this scale as a measure of scholastic aptitude is to be compared with a correlation of .39 between grade point averages and the scores on the American Council on Education *Psychological Examination* for the validating group (N = 81).

A correlation was found between the scores made with the Group II key (composed of mathematicians, engineers, chemists, and physicists) (17) of the *Strong Vocational Interest Blank* and the scores of the "theoretical" key. A correlation of .80 was obtained. This high correlation would seem to indicate that these two keys are to a large extent measuring approximately the same thing.

In order to determine whether the weightings on this new "economic" scale were similar to the Strong Key for Group VIII (accountant, office man, purchasing agent, banker) (17) a correlation was found for the scores of 89 subjects on the two keys. A correlation of .21 was obtained from these data, thus indicating that there is apparently little relationship between these keys. It would seem that the "economic" man is a different type of individual from the business man as represented by the accountant, office man, purchasing agent, banker group. The recent factor analysis study made by Ferguson, Humphreys, and Strong (11) shows how eight of the Strong scales (teacher, life insurance salesman, certified public accountant, office worker, physician, lawyer, Y. M. C. A. secretary, and chemist) are related to the six Allport-Vernon scales. The results indicate that the Strong scales used contain small factor loadings in the fifth factor. It is quite likely that the Strong "economic" scale of this study may produce heavier loadings and be an additional help for counseling in this area.

It seemed possible that a better estimation of the value of the "social adjustment" scale could be made by comparing resulting scores with social activity records. The 10 students of the validating group scoring highest on the new Strong "social adjustment" scale were compared with the ten students scoring lowest. A count was made of the number of activities in which these students participated during the semester the tests were taken. The students composing the "social" group belonged to a total of seven organizations, while those making up the "non-social" group belonged to only two.

Summary and Conclusion

The purpose of this study was to investigate the possibility of using the *Strong Vocational Interest Blank* to measure certain personality traits previously measurable only by the use of several other tests.

An analysis of the results showed that all of the scores based on the new keys were positively correlated with the traits measured. Validation coefficients based on a new group of subjects were: Bernreuter F1-C, .48; Bernreuter F2-S, .52; Allport-Vernon Theoretical, .48; Allport-Vernon Economic, .56; Bell Home Adjustment, .21; Bell Health Adjustment, .34; Bell Social Adjustment, .50; Bell Emotional Adjustment, .07; and the American Council on Education, Intelligence, .45. The reliabilities of these keys ranged from .70 for the Health Adjustment key to .87 for the Social Adjustment key.

In presenting the findings of this investigation, it should be remembered that the items of the *Strong Vocational Interest Blank* were not designed to be used as elements of a personality test. However, the weighted items of the Strong Blank give a fairly good picture of some of the factors making up a particular trait, e.g., Spranger's "economic" man (as measured by the Allport-Vernon test) and the weighted items on the new Strong "economic" key.

From the data gathered, the reliabilities indicate that the new Strong keys are fairly consistent in the material they are

measuring. It also appears that some traits can be measured with more accuracy by the Strong Blank than other traits. It does not seem advisable to continue work in the fields of "home adjustment", "health adjustment", or "emotional adjustment" with the *Strong Vocational Interest Blank* because of low validities obtained in these fields.

From this study, however, it does appear that a prediction of self-confidence and of sociability can be made with a fair amount of accuracy with the Strong Blank, and that types of individuals such as "theoretical" and "economic" (business man) can be determined fairly well by using the Strong keys for measurement of these traits. "Intelligence" scores based on responses to the Strong Blank show as high a correlation with college success as the American Council on Education *Psychological Examination* scores. This would indicate that in spite of relatively low validity coefficients when correlated with scores on the original tests, others of the new scales may be found to possess considerable practical validity in the evaluation of socially significant behavior.

REFERENCES

1. Allport, F. H. *Social Psychology.* New York: Houghton Mifflin Company, 1924. Chap. V, VI, XIV.
2. Allport, G. W. "A Test for Ascendance-Submission", *Journal of Abnormal and Social Psychology,* XXIII, (1928), 118-136.
3. Allport, G. W. "Concepts of Trait and Personality", *Psychological Bulletin,* XXIV, (1927), 284-293.
4. Allport, G. W. *Personality.* New York: Henry Holt and Company, 1937. pp. 236-237; 303-304.
5. Allport, G. W. and Vernon, P. E. *Manual of Directions for a Study of Values.* Chicago: Houghton Mifflin Company, (1931).
6. *Manual for American Council on Education Psychological Examination.* Washington: American Council on Education, (1940).

7. Bell, Hugh M. *Manual for The Adjustment Inventory,* Stanford University Press, California, (1934).

8. Bernreuter, R. G. *Manual for The Personality Inventory,* Stanford University Press, California, (1938).

9. Chambers, O. R. "Character Trait Tests and Prognosis of College Achievement", *Journal of Abnormal and Social Psychology,* XX, (1925), 303-311.

10. Conklin, E. S. "Three Diagnostic Scorings for the Thurstone *Personality Schedule",* Indiana University Publications, Science Series, No. 6, (1937).

11. Ferguson, L. W., Humphreys, L. G., and Strong, F. W. "A Factorial Analysis of Interests and Values", *Journal of Educational Psychology,* XXXII (1941), 197-204.

12. Kelley, T. L. "The Scoring of Alternative Responses with Reference to Some Criterion", *Journal of Educational Psychology,* XXV, (1934), 504-510.

13. Kelly, E. L., Terman, L. M., and Miles, C. C. "Ability to Influence One's Score on a Typical Paper and Pencil Test of Personality", *Character and Personality,* IV, (1936), 206-215.

14. Kelly, E. L. "A Preliminary Analysis of Psychological Factors in Assortative Mating", *Psychological Bulletin,* XXXIV, (1937), 749.

15. Otis, A. S. *Manual for Self-Administering Tests on Mental Ability,* Yonkers-on-Hudson, New York: World Book Company, (1922).

16. Steinmetz, H. C. "Measuring Ability to Fake Occupational Interest", *Journal of Applied Psychology,* XVI, (1932), 123-130.

17. Strong, E. K., Jr. *Manual for Vocational Interest Blank for Men,* Stanford University Press, California, (1940).

18. Strong, E. K., Jr. "Procedure for Scoring an Interest Test", *Psychological Clinic,* XIX, (1930), 63-72.

19. Strong, E. K., Jr. "Interest Maturity", *Personnel Journal,* XII, (1933), 77-90.

20. Strong, E. K., Jr. "Interests of Men and Women", *Journal of Social Psychology,* VII, (1936), 49-67.

21. Young, C. W. and Estabrooks, G. H. "Reports on the Young-Estabrooks 'Studiousness Scale' for Use with Strong Vocational Interest Blank for Men", *Journal of Educational Psychology,* XXVIII, (1937), 176-187.

A STUDY OF THE GENTRY VOCATIONAL INVENTORY

CLIFFORD FROEHLICH
State of North Dakota Occupational Information and Guidance Service

THE *Vocational Inventory* developed by Curtis G. Gentry, Director of Guidance and Secondary Education, Public Schools, Knoxville, Tennessee, contains 434 questions.[1] Three hundred and eighty-four are in the *Inventory* proper, and the remainder constitute a personality inventory. The *Inventory* proper, according to Gentry's statement, "classifies the applicant's strengths and weaknesses with reference to these eight major groups"[2]: (1) social service, (2) literary work, (3) business, (4) law and government, (5) art, (6) mechanical designing, (7) mechanical construction, and (8) science. Gentry states in the *Manual of Directions* that the *Inventory* was constructed after "a fruitless search for an instrument which would yield a general vocational over-view of pupils and young adults."[2] The *Inventory* began to take shape in 1921 and has undergone many revisions since that time. It was copyrighted and published in its present form in 1940.

This study of the *Vocational Inventory* is based upon results obtained by administering it to 815 seniors in the high schools of Cass County, North Dakota. A majority (72 per cent) of the students were from a single school located in the metropolitan area of the county. In all cases the test was

[1] C. G. Gentry. *Vocational Inventory,* (Nashville: Educational Test Bureau, 1940).

[2] C. G. Gentry. *Manual of Directions, Vocational Inventory,* (Nashville: Educational Test Bureau, 1941).

administered by persons specially trained in tests and measurements and was given under favorable testing conditions.

A statistical summary of scores of 1,000 Knoxville High School seniors is reported in the manual. Table 1 shows the total number of these students who received the highest scores in each of the groups as well as the percentages as reported by Gentry. The possible range of scores is from zero to 180. The highest score made by each student indicates the field for which greatest interest and ability are possessed.

TABLE 1

NUMBER OF HIGH SCORES BY GROUPS AS REPORTED BY GENTRY[3]

GROUP	I	II	III	IV	V	VI	VII	VIII
Total	168	135	143	184	99	23	172	76
Percent	16.8	13.5	14.3	18.4	9.9	2.3	17.2	7.6

A similar tabulation was made of the 815 students that were tested in Cass County. Table 2 gives the results of this tabulation. Separate tabulations are given by sexes.

TABLE 2

NUMBER OF HIGH SCORES BY GROUPS, 815 CASS COUNTY SENIORS

GROUP	I	II	III	IV	V	VI	VII	VIII
Girls Total	60	33	41	123	60	3	0	85
Percent	14.8	8.2	10.	30.4	14.8	0.7	0	21.
Boys Total	5	5	42	42	13	53	179	71
Percent	1.2	1.2	10.3	10.3	3.2	12.7	43.7	17.2
Total	65	38	83	165	73	56	179	156
Percent	7.9	4.6	10.2	20.2	8.9	6.9	21.9	19.4

It is evident from a comparison of girls and boys in Table 2 that there is a significant sex difference which must be taken into account in interpreting the results of this test although Gentry makes no statement in the *Manual* regarding sex differences. It is also interesting to note that when the scores of boys and girls are combined, there is still a wide discrepancy between the percentages as reported by Gentry and the percentages as found in the present study.

The question may be raised as to whether separate norms should be established for each sex in instances where sex differences are as pronounced as this study reveals. There is a

[3]Ibid, p. 8.

considerable difference of opinion on this point. Those who maintain that combined norms are justified base their opinion on the fact that the type and number of jobs open to women differs materially from those open to men. To them, establishing separate norms presumes that as many women as men would be advised to go into the field of mechanical construction, for example, when it is evident that many more men than women should be so advised. However, this position is not tenable for the writer. The establishment of separate norms would not necessarily mean that as many women as men would be guided into the mechanical construction field, but it would emphasize to the users of the inventory that sex differences do exist and that they must be taken into account. No instrument of this type can in any way supplant the necessity for every counselor's having a knowledge of the job demands and opportunities for both men and women in all occupational categories.

Question 383 of the *Inventory* asks the student to "name three vocations which appeal to you at the present time. Name your first choice first, if you have a first choice."[4] The writer noted that while using this instrument in a clinical situation there seemed to be a large proportion of the students whose claimed first choice of an occupation was in the same group as their highest score on the test. In order to check this relationship statistically, the student's personal choice was classified into one of the eight groups according to the "Psychological Classification of Occupations" as given by Gentry. The contingency coefficient for student's first choice and the highest score on the *Inventory* for 338 female high school seniors is .718. The *C* on a similar comparison for 300 male high school seniors is .733. When these two groups are combined, *C* equals .751. This correlation raises the question of whether the inventory, in a large measure, may not simply be a re-statement of the student's personal choice. An examination of the scattergrams for the correlations given above

[4]C. G. Gentry. *Op. cit.* p. 27.

indicates that 51 per cent of the cases fall on the axis, indicating that 51 per cent of the high school seniors have a personal vocational choice that is in the field of their highest score. Darley[5] reports a maximum contingency coefficient of .5 between the claimed occupational interest types and measured occupational interest patterns for approximately 1,000 cases, with the Strong test, compared to .751 revealed by this study. Darley felt that he obtained a contingency coefficient as high as .5 only because he forced both the claimed and measured interest types into broad categories.

This close agreement between the scores on the *Inventory* and the student's stated choice may merely mean that he has, by a long and thorough process, come to such complete knowledge of occupational requirements that he reached the same conclusion that the inventory revealed in a comparatively short time. However, the individuals upon whom this study is based were members of school populations where no organized effort had been made to provide occupational information, and only a few of the students had related work experience. Furthermore, the counselors who interviewed the students following the administration of the *Inventory* found that in the majority of the cases the students had a very meager knowledge of the requirements of their stated vocational choices.

In the *Manual,* Gentry states, "An analysis of the *Inventory* returns on this sampling indicates that the second highest score is usually earned in an occupational group related to the first."[6] Tables 3, 4, and 5 show the relationship of the highest score to the second highest score.

In some cases, Gentry's statement seems to be correct. However, in many cases it is difficult to find justification for his statement. For example, Group IV (Business), as shown in Table 5, indicates that 50 people earned their second score in Group III (Law and Government); 22 in Group II (Lit-

[5]J. G. Darley. *Clinical Aspects and Interpretation of the Strong Vocational Interest Blank* (New York: The Psychological Corporation, 1941).

[6]C. G. Gentry. *Op. cit.* p. 4.

TABLE 3

RELATIONSHIP OF HIGHEST TO SECOND HIGHEST SCORE

328 High School Females

Group in which student attained highest score

	Group	I	II	III	IV	V	VI	VII	VIII	Total
Group	I	0	8	5	17	7	0	0	25	62
in which	II	9	0	5	20	9	0	0	2	45
student	III	6	14	0	36	10	0	0	10	76
attained	IV	12	4	12	0	15	0	0	21	64
second	V	4	3	2	13	0	1	0	5	28
highest	VI	0	0	0	1	7	0	0	1	9
score	VII	0	0	0	0	0	0	0	2	2
	VIII	9	3	14	16	0	0	0	0	42
	TOTAL	40	32	38	103	48	1	0	66	

TABLE 4

RELATIONSHIP OF HIGHEST TO SECOND HIGHEST SCORE

289 High School Males

Group in which student attained highest score

	Group	I	II	III	IV	V	VI	VII	VIII	Total
Group	I	0	0	1	0	0	0	0	1	2
in which	II	0	0	2	2	0	0	0	1	5
student	III	0	2	0	14	3	2	4	7	32
attained	IV	0	0	10	0	1	2	13	4	30
second	V	0	2	0	0	0	1	2	2	7
highest	VI	1	0	0	2	3	0	59	7	72
score	VII	1	0	10	12	2	36	0	28	89
	VIII	1	0	3	6	2	6	34	0	52
	TOTAL	3	4	26	36	11	47	112	50	

TABLE 5

RELATIONSHIP OF HIGHEST TO SECOND HIGHEST SCORE

617 Males and Females Combined

Group in which student attained highest score

	Group	I	II	III	IV	V	VI	VII	VIII	Total
Group	I	0	8	6	17	7	0	0	26	64
in which	II	9	0	7	22	9	0	0	3	50
student	III	6	16	0	50	13	2	4	17	108
attained	IV	12	4	22	0	16	2	13	25	94
second	V	4	5	2	13	0	2	2	7	35
highest	VI	1	0	0	3	10	0	59	8	81
score	VII	1	0	10	12	2	36	0	30	91
	VIII	10	3	17	22	2	6	34	0	94
	TOTAL	43	36	64	139	59	48	112	116	

erary); 22 in Group VIII (Science); 17 in Group I (Social Service); 13 in Group V (Art); and 12 in Group VII (Mechanical Construction). It appears that in 64 per cent of the cases the second highest scores will fall in some other group besides Group III. This is certainly a wider distribution than is indicated by the statement "usually earned in an occupational group related to the first."

The inference of Gentry's statement, it would seem, is that there are fairly high intercorrelations among certain scores if the second score is related to the first. Such overlapping, of course, is undesirable from a standpoint of measurement efficiency. The finding that such relationship is low therefore is a point in favor of the blank, even though it discredits Gentry's statement of close relationship.

In the *Manual of Directions,* Gentry assumes that interest will drive the person to acquire certain objective information or proficiency in the field of interest. According to Fryer's[7] summary of the literature in the field, this so-called objective approach to the measurement of interest by the way of achievement or proficiency has never been unusually successful.

According to the statement of Gentry, the test is designed to measure the student's "strengths and weaknesses." Assuming that ability to deal with linguistic concepts and achievement in English are necessary requisites for success in the literary field, the following study was made. The Literary Group scores of 176 high school senior girls, selected at random, were compared with the scores made on the *Cooperative English Test* Form O.M., and the L-score of the American Council on Education *Psychological Examination,* 1938 Edition. The correlation between the English achievement score and the L-score was .73, with a P.E. of .022. A correlation of .74 with a P.E. of .018 was obtained by Beyers[8] for these same measures on 500 college freshmen. The correlation between English achievement and the Literary Key on the Gentry *Inventory* was .589 with a P.E. of .033. The correlation was .605, with a P.E. of .055 between the L-score and the Literary Key. A partial correlation with the Gentry score held constant was .58. With the L-score held constant the correlation between the Literary Key and English achievement

[7]Douglas Fryer. *The Measurement of Interest* (New York: Henry Holt and Company, 1931).

[8]Otto Beyers. "Report of the Freshmen Testing Program, N. D. A. C., Fargo, North Dakota", 1940.

was .095; with the English score held constant the correlation between the Literary Key and the L-score was .297.

These correlations seem to indicate that, to some extent, the Gentry test is measuring some factor which is not measured by the English and intelligence examinations. There is a question as to just what is being measured; possibly the high relationship between the personal choice and measured choice may offer one explanation. A common factor of experience in an academic high school may be another factor.

When intelligence is held constant, there seems to be little relationship between the English and the Gentry scores for the Literary Key. It is evident that the Literary Key gets at very little related to English achievement which is not already measured by an intelligence test. Yet in The Student's Manual that accompanies the test, the author states concerning this group that "one entering an occupation in the above group should like literature and English. He (or she) should be very much interested in composition, in writing articles, poems, themes, and reports."[9]

In making the above analysis to test the relationship of a single key to known objective measures, the writer does not contend that an interest inventory should show a marked positive relationship with such measures. On the contrary, an interest inventory can hardly be justified if it adds nothing to the scores of other measures. Though the findings here do not entirely discredit the usefulness of the *Vocational Inventory* as a measure of interest in the literary area, it does tend to disprove Gentry's claim that he is measuring "strengths and weaknesses."

While this study does not present conclusive evidence, the results at least indicate that this test should be more carefully standardized and evaluated before it is used in a counseling situation. The writer feels that the scores it now yields are of questionable value to the average guidance worker. Refine-

[9]C. G. Gentry. *Individual Analysis Report.* (Nashville: Educational Test Bureau, p. 6, 1940).

ment of procedures and further occupational standardization and follow-up data are necessary before the counselor can regard this inventory as a valid diagnostic tool.

THE RELATIONSHIP OF THE AFFECTIVE TOLERANCE INVENTORY TO OTHER PERSONALITY INVENTORIES

ROBERT I. WATSON
College of the City of New York

IN PRESENTING any new personality inventory it is important to establish the nature and extent of relationship that it bears to other scales designed to measure similar aspects of personality. The following pages present data on the relationship of the Watson-Fisher *Inventory of Affective Tolerance* to other standardiized measures.[1]

In constructing most inventories of emotional stability, items successfully used by others were usually selected from the clinical literature and combined with additional items for preliminary tryout with a more or less pragmatic outlook. If they "hung together" on internal validation and differentiated between extreme groups, a new inventory was constructed.

The *Inventory of Affective Tolerance* originated from a somewhat more theoretical framework. Items were collected with certain theoretical predilections as guiding principles. Statements of "symptom" were included in the preliminary tryout only if they seemed to be appropriate to this trait of affective tolerance. As a result, many so-called conventional items were not included. A brief description of this point of view follows.

An individual's affective tolerance is judged to be his ca-

[1]This present report is considered as preliminary to a study by means of some factor technique.

pacity to handle his affective tensions; his capacity to adjust to affective disturbances. Among the chief aspects of affective tolerance are the capacities to withstand or endure emotional tension, to vent or discharge emotional tension, and to govern or direct emotional tension. It is the aggregate of these that the *Inventory of Affective Tolerance* purports to measure. A more complete discussion, including a description of the validation of the inventory, is given in an article by Fisher and Watson (3).

The Relationship to the Watson-Fisher Inventory of Affective Potency, *to the Willoughby* Personality Schedule, *and to the Bernreuter* Personal Inventory

The subjects employed in this section of the inquiry consisted of 55 boys and 97 girls, students at the University of Idaho, Southern Branch, who were described in connection with previous papers (3) (10). Besides the *Inventory of Affective Tolerance,* certain other measures including measures of general aptitude,[2] the Watson-Fisher *Inventory of Affective Potency,* the Willoughby *Personality Schedule,* and the Bernreuter *Personality Inventory* were administered to these students during the same semester.

The *Inventory of Affective Potency* purports to measure the trait of affective arousability, the strength and duration of our everyday affective responses (10). The Willoughby *Personality Schedule* is designed to measure "neurotic tendencies" (11). The Bernreuter *Personality Inventory* is designed to measure several aspects of personality. In view of the analyses of Flanagan (4) and Lorge (8), attention will be given to but two of these measures, designated by the symbols F1-C and F2-S in the inventory manual (2). The first is a measure of confidence in oneself. Persons scoring

[2]The correlations of the *Inventory of Affective Tolerance* with general aptitude were negligible. With the Otis *Self-Administering Test of Mental Ability, Higher Examinations,* the correlation coefficients were respectively, —.04 and —.02 for 50 boys and 90 girls. The correlations with the American Council on Education *Psychological Examination,* 1939 edition, were respectively, .04, .03, and .04 for L, Q, and Total scores in the case of the boys, and .01, .10, and .06 for the girls.

low tend to be self-confident and well adjusted, and individuals scoring high tend to have feelings of inferiority. The second of these measures is one of sociability. At one extreme individuals tend to be non-social, at the other, gregarious.

TABLE 1

THE CORRELATION BETWEEN THE INVENTORY OF AFFECTIVE TOLERANCE AND THE WATSON-FISHER INVENTORY OF AFFECTIVE POTENCY, THE WILLOUGHBY PERSONALITY SCHEDULE, AND THE BERNREUTER PERSONALITY INVENTORY

	Tolerance					
	Boys			Girls		
	N	r	k	N	r	k
Watson-Fisher	55	—.14	.99	97	—.21	.98
Willoughby	47	—.70	.71	94	—.69	.73
Bernreuter F1-C.....	46	—.66	.75	87	—.56	.83
Bernreuter F2-S.....	46	—.13	.99	87	—.11	.99

Table 1 gives the correlation between the scores on these inventories and scores on the tolerance inventory together with the corresponding coefficients of alienation. The correlations with the Willoughby and the F1-C or confidence factor of the Bernreuter test are substantial, ranging between —.56 and —.70. The correlations considerably exceed the minimum demanded at the one per cent level of significance (7). The correlations with the *Inventory of Affective Potency* and the sociability factor, F2-S, of the Bernreuter test are not significant at the one per cent level.

It would appear, then, that affective tolerance, as herein described and measured, bears considerable relation to self-confidence and lack of neurotic tendency, but little relation to sociability or affective potency.

This substantial relationship, however, should not be interpreted to mean that the tolerance inventory is so closely related to these other scales as to be superfluous for the measurement of individual differences. Consideration of the coefficients of alienation is pertinent in this connection. Reduction of the standard error of estimate by one-half, as expressed by a k of .50, requires that r be .866. The smallest coeffi-

cient of alienation found here is .71. Garrett (5) states that "For *r*'s of .80 or less the coefficients of alienation are clearly so large that predictions of individual scores based upon the regression equation are little better than a 'guess'." Although a substantial relation does exist between scores on the affective tolerance inventory and scores on these other inventories, it is evident that the *Inventory of Affective Tolerance* measures something other than whatever is measured by these two inventories.

The Relationship to the Colgate Personal Inventory C 2 *and the Bell* Adjustment Inventory

Fifty-nine white female student nurses, who were tested while receiving three months of their training at the Idaho State Mental Hospital (South) at Blackfoot, Idaho, took the *Inventory of Affective Tolerance,* the *Colgate Personal Inventory C2,* and the Bell *Adjustment Inventory.*[3] The background of these subjects is summarized in Table 2.

TABLE 2

DESCRIPTION OF THE BACKGROUND OF FIFTY-NINE STUDENT NURSES

	Mean	σ	Range
Age	23.25	2.56	19-33
School Year Completed	12.61	1.09	12-16
Months of Training Completed	28.87	4.50	12-33
Otis *Self-Administering Test* Scores	45.90	9.18	25-65
Moss *Nursing Aptitude Test* Scores	139.90	30.02	41-186

The *Colgate Personal Inventory C 2* is designed to measure traits of introversion-extroversion. (6) The Bell *Adjustment Inventory* measures home, health, social, emotional, and occupational adjustment.[1] The higher the score in this inventory, the more unsatisfactory the adjustment.

The product-moment correlations between the *Inventory of Affective Tolerance* and the other inventories are presented in Table 3. All correlations are negative and all but two are significant at the one per cent level. (7) The correlation with the Colgate *Personal Inventory C 2* is not quite

[3] I wish to acknowledge the courtesy of Mr. Barney Bybee, then psychometrician at the Idaho State Mental Hospital, South, in supplying these scores.

TABLE 3

THE CORRELATIONS BETWEEN THE INVENTORY OF AFFECTIVE TOLERANCE AND THE COLGATE PERSONAL INVENTORY C 2 AND THE BELL ADJUSTMENT INVENTORY

Measure	N	Tolerance	
		r	*k*
Colgate C-2	59	—.30	.95
Bell Home	59	—.40	.92
Bell Health	59	—.11	.99
Bell Social	59	—.60	.80
Bell Emotional	59	—.54	.84
Bell Vocational	45	—.50	.87
Bell Total[1]	59	—.56	.83

[1]Vocational items are omitted since not all subjects took the form of the Bell Inventory that includes this section.

significant, and the correlation with the Bell Health Score is entirely negligible. Health adjustment and introversion-extroversion are apparently not significantly correlated with affective tolerance. The remaining correlations with Bell scores range from —.40 to —.60. Affective tolerance bears some relation to home, social, emotional, vocational, and total adjustment as measured by the Bell Inventory.

Although there is evidence that some degree of relationship exists, nevertheless the coefficients of alienation, which are also reported in Table 3, do not encourage the view that the two inventories can be used interchangeably. The smallest coefficient of alienation is .80, which implies 20 per cent efficiency in prediction.

Commonality of Items

An attempt was made to find out what items the Watson-Fisher inventory and the other personality measures had in common. Such an inquirty is pertinent in a field of investigation marked by repetition of items from inventory to inventory. Does the approach from the theoretical point of view earlier expressed result in the selection of the same items as the more pragmatic approach?

Two degrees of similarity of content can be distinguished. The first category adopted included items in which the con-

tent apparently did not differ in scope, e. g., item 4 of the tolerance blank, *"I keep in the background at social gatherings"*, as compared to item 118 of the Bernreuter, *"Do you keep in the background at social functions?"* In the second category were items in which the similarity was one of a whole to a part or a part to a whole, e. g., item 60 of the tolerance inventory, *"I control my feelings of grief or sorrow"*, as compared to item 35 of the Bell, *"Are you easily moved to tears?"* The two categories will be referred to as "similar" and "partially similar," respectively.

The items of the tolerance inventory were checked against the four other personality measures used in one or another of the two samples studied for similar and partially similar items. The results of this subjective, somewhat crude analysis are presented in Table 4.

TABLE 4

THE SIMILARITY OF CONTENT OF ITEMS OF CERTAIN PERSONALITY MEASURES TO THE 61 ITEMS CONTAINED IN THE INVENTORY OF AFFECTIVE TOLERANCE

Measure	Similar		Partially Similar		Similar or Partially Similar	
	N	%	N	%	N	%
Bernreuter	16	26	2	3	18	30
Willoughby	7	11	5	8	12	20
Laird C-2	6	10	4	7	10	16
Bell	14	23	4	7	18	30

It is apparent that similarity of content expressed in this fashion does occur. However, not more than 30 per cent of the items of the *Inventory of Affective Tolerance* appear in any one of these measures.

The commonality of items can be expressed in another fashion, namely, the total number of the 61 tolerance items appearing in one or more of the other measures as either a "similar" or a "partially similar" item. There were 27 such items. It is evident, then, that many of the items reported as similar in the data of Table 4 were found in two or more

of the other blanks. In all, 34 items or 56 per cent of the items in the tolerance inventory do not appear in the other four blanks. Apparently, then, there is no great similarity of items in the tolerance inventory and those contained in the other inventories studied.

Conclusions

1. Affective tolerance bears substantial relationship to confidence in oneself, lack of neurotic tendency, and social and emotional stability as measured by the scales used.

2. Little or no relation is found between affective tolerance and affective potency, sociability, or health adjustment as measured by the scales used.

3. No relations are found to be of such a magnitude as to make the *Inventory of Affective Tolerance* superfluous, since the smallest coefficient of alienation is .71.

4. More than half the items contained in the *Inventory of Affective Tolerance* do not appear in the other personality measures employed.

REFERENCES

1. Bell, H. M. *Manual for the Adjustment Inventory: Adult Form.* Stanford University: Stanford University Press, 4 pp.
2. Bernreuter, R. G. *Manual for the Personality Inventory.* Stanford University: Stanford University Press, 6 pp.
3. Fisher, V. E. and Watson, R. I. "An Inventory of Affective Tolerance", *Journal of Psychology,* XII (1941), 149-157.
4. Flanagan, J. C. *Factor Analysis in the Study of Personality.* Stanford University: Stanford University Press, 1935. 103 pp.
5. Garrett, H. E. *Statistics in Psychology and Education.* New York: Longmans, 1937. 493 pp.
6. Laird, D. A. *General Information and Directions for Using the Colgate Tests of Emotional Outlets.* Hamilton: Hamilton Republican, 4 pp.
7. Lindquist, E. F. *Statistical Analysis in Educational Research.* Boston: Houghton-Mifflin, 1940. 266 pp.

8. Lorge, I. "Personality Tests by Fiat: I The Analysis of the Total Trait Scores and Keys of the Bernreuter Personality Inventory", *Journal of Educational Psychology,* XXVI (1935), 273-278.

9. Otis, A. S. *Manual of Directions and Key: Otis Self-Administering Tests of Mental Ability*. Yonkers-on-Hudson: World Book, 12 pp.

10. Watson, R. I. and Fisher, V. E. "An Inventory of Affective Potency", *Journal of Psychology,* XII (1941), 139-148.

11. Willoughby, R. R. *Directions: Willoughby (Clark-Thurstone) Personality Schedule.* Providence: Author, 2 pp.

THE INFLUENCE OF TRAINING ON MECHANICAL APTITUDE TEST SCORES

RICHARD W. FAUBION and EARLE A. CLEVELAND
Air Corps Technical Training Command

and

THOMAS W. HARRELL
University of Illinois

THE present study is designed to investigate the influence of training on scores of the *Mechanical Movements* and *Surface Development Tests* (similar to many so-called aptitude tests), which look highly susceptible to training. In previous studies Harrell and Faubion found the *Mechanical Movements* and *Surface Development Tests* to be valid in predicting course grades of student airplane mechanics. The *Surface Development Test* correlated .55 with composite airplane mechanics grades for 84 men (1), and .47 with a slightly different composite for 105 other men (2). For the same two groups the correlations of the *Surface Development Test* with mechanical drafting and blueprint reading were .54 and .50, respectively. The *Mechanical Movements Test* correlated .39 and .26 with composite grades in the same two groups.

These tests developed by Thurstone are fairly familiar (3). The *Mechanical Movements Test* requires an individual to figure out how machine parts, particularly gears and pulleys, work. The *Surface Development Test* involves matching similar parts for drawings shown in two dimensions and in three-dimensional perspective.

The procedure was to study two groups which were matched for mental test scores but which differed in the

amount of mechanical training they had received. The two groups were each composed of 100 soldiers and were matched on the basis of scores on a mental test similar to the Henmon-Nelson. One group consisted of Air Corps recruits who had not as yet been entered in any of the technical courses offered by the Air Corps Technical Schools. The second group was composed of airplane mechanics students who had just finished a six-week basic training course in mechanical drafting and blueprint reading, elements of metalwork, elements of electricity, shop mathematics, and air corps fundamentals. The mean raw score for the untrained recruits on the mental test was 57.0, as compared with a mean raw score of 57.3 for the trained group. The difference between the two means is only one-fourth the standard error of the difference and consequently is quite insignificant. The variability of the two groups is also similar, a standard deviation of 8.96 being found for the recruits as against a standard deviation of 8.97 for the students.

The two groups were not intentionally paired for previous mechanical experience or training, but were selected at random from groups of Air Corps recruits and Air Corps Technical School students, respectively.

Special attention will be given to the mechanical drafting and blueprint reading course, as well as the elements of metalwork course, because of their apparent similarity to the tests. Mechanical drafting and blueprint reading, a forty-hour course, had the following outline:

a. Fundamental principles of mechanical drafting.
b. Exercises in orthographic projection.
c. Development of surfaces.
d. Blueprint reading.
e. Exercises in blueprint reading.

Elements of metalwork, a sixty-hour course, had the following outline:

a. Properties and uses of the common metals.

b. The care and use of the common tools needed in the repair and manufacture of small parts.
c. Metalwork—projects in drilling, filing, thread cutting, reaming, etc.
d. Soldering—soft and hard soldering.
e. Brazing.

The results given in Table 1 show no significant differences. The trained men had a score of 18.4 on the *Surface Development Test* as compared with 17.9 for the recruits. This difference is less than half the standard error of the difference.

TABLE 1

COMPARISON BETWEEN 100 RECRUITS AND 100 SOLDIERS TRAINED IN BASIC AIRPLANE MECHANICS

	Mean for Recruits	Mean for Trained Men	Difference	Sigma of Difference
Mental Test	57.0	57.3	0.3	1.2
Surface Development	17.9	18.4	0.5	0.9
Mechanical Movements (No. Right)	32.2	32.5	0.3	1.2
Mechanical Movements (Rt. —W)	18.2	17.6	0.6	1.7

The *Mechanical Movements Test* scores were treated in two ways: (1) Number right and (2) Rights minus wrongs. Since the number of choices varies from one question to another, the use of a simple correction formula would be questionable. The trained group had a mean score of .3 of a point higher for the number right. The recruits had a mean score .6 of a point higher when the correction formula was used. Neither of these differences is as large as half the standard error of the difference.

These results indicate that six weeks of intensive training in mechanical courses do not significantly increase mechanical aptitude test scores, even where the test is very similar to the activities carried out in the training. This is strikingly true of the *Surface Development Test,* in which the items resemble mechanical drafting and blueprint reading work. No conclusion can be drawn as to how far this result can be generalized; possibly longer training or earlier training would show a

significant increase. The authors wish to point out, however, that the present results are contrary to statements often made about a mechanical aptitude test.

REFERENCES

1. Harrell, T. W. and Faubion, R. W. "Selection Tests for Aviation Mechanics", *Journal of Consulting Psychology,* IV, (1940), 104-105.
2. ———— "Primary Mental Abilities and Aviation Maintenance Courses", *Educational and Psychological Measurement,* I, (1941), 59-66.
3. Thurstone, L. L. *Primary Mental Abilities.* Psychometric Monographs, No. 1. Chicago: University of Chicago Press, 1938. 128pp.

NEW TESTS*

American School Achievement Tests, by Robert V. Young, Willis E. Pratt, and Frank Gatto. 1941. Forms A and B. Primary Battery I, for grade 1. Time, 35 minutes. $2.50 per 100; 3c each; specimen set 25c. Primary Battery II, for grades 2 and 3. Time, 85 minutes. $4.00 per 100; 5c each; specimen set 30c. Published by the Public School Publishing Company, 509-513 North East Street, Bloomington, Illinois.

American School Reading Readiness Test, by Robert V. Young, Willis E. Pratt, and Carroll A. Whitmer. For kindergarten and grade 1. Time, about 30 minutes. Form A, $4.00 per 100; 5c each; specimen set 25c. Published by the Public School Publishing Company, 509-513 North East Street, Bloomington, Illinois.

Arithmetical Reasoning Test, by Alfred J. Cardall. 1941. For academic and technical prediction. For 12th grade level and above. Time, 40 minutes. Forms A and B, 5c each; specimen set 20c. Published by Science Research Associates, 1700 Prairie Avenue, Chicago, Illinois.

Chicago Tests for Primary Mental Abilities, by L. L. Thurstone and Thelma Gwinn Thurstone. 1942. For ages 11 to 17. Time, 40 minutes for each of six booklets. $5.00 per 25 sets; $9.00 per 50 sets; $15.00 per 100 sets; $70.00 per 500 sets; specimen set $1.00; supplementary supplies additional. Published by the American Council on Education, 744 Jackson Place, Washington, D. C.

College English Test, National Achievement Tests, by A. C. Jordan. 1941. For high school seniors and college freshmen. Time, about 45 minutes. Forms A and B, $2.50 per 25; 100 or more copies 7½c each. Published by the Acorn Publishing Company, Rockville Centre, Long Island, New York.

---

*Prepared by Jane Gilbert.

Detroit Alpha Intelligence Test, by Harry J. Baker. 1941. For grades 4 to 8. Time, 32 minutes. Forms S and T, $3.50 per 100; 4c each; specimen set 15c. Published by the Public School Publishing Company, 509-513 North East Street, Bloomington, Illinois.

English Minimum Essentials Test, by J. C. Tressler. Revised 1941. For grades 8 to 12. Time, about 40 minutes. Forms A, B, and C, 75c per 25; 4c each; specimen set 10c. Published by the Public School Publishing Company, 509-513 North East Street, Bloomington, Illinois.

Every-Day Life, by Leland H. Stott. 1941. To measure three factors in self-reliance. For high school students. Time, about 30 minutes. Hand- or machine-scored. $4.00 per 100; $2.25 per 50; specimen set 15c; machine-scoring answer sheets, $2.00 per 100 up to 500. Published by the Sheridan Supply Company, P. O. Box 837, Beverly Hills, California.

Furbay-Schrammel Social Comprehension Test, by John H. Furbay and H. E. Schrammel. 1941. For high school and college students, and adults. Time, 80 minutes. Form A, $1.70 per 25; 7c each; specimen set 15c. Published by the Bureau of Educational Measurements, Kansas State Teachers College, Emporia, Kansas.

Interest Inventory for Elementary Grades, by Mitchell Dreese and Elizabeth Mooney. 1941. For grades 4, 5, and 6. Time, about 30 minutes. Published by Center for Psychological Service, George Washington University, Washington, D. C.

Inventory of Social Behavior, by Ellis Weitzman. 1941. For ages 16 to 25. Time, about 20 minutes. $4.00 per 100; $2.25 per 50; specimen set 15c. Published by the Sheridan Supply Company, P. O. Box 837, Beverly Hills, California.

Iowa Every-Pupil Tests of Basic Skills—Form M, by H. F. Spitzer, Ernest Horn, Maude McBroom, H. A. Greene, and E. F. Lindquist. 1941. Test A, Silent Reading Comprehension; Test B, Work-Study Skills; Test C, Basic

Language Skills; Test D, Basic Arithmetic Skills. Elementary Battery for grades 3 to 5. Time, about 85 minutes. $1.15 per 25 for each test; $3.75 per 25 for complete battery. Advanced Battery for grades 5 to 8. $1.25 per 25 for each test; $4.00 per 25 for complete battery. Published by Houghton Mifflin Company, 2 Park Street, Boston, Massachusetts.

Iowa Placement Examinations, English Training—Form M, constructed by M. F. Carpenter, G. D. Stoddard, and L. W. Miller; revised by M. F. Carpenter and D. B. Stuit. Revised 1941. For college students. Time, 45 minutes. Hand- and machine-scored. $4.00 per 100; machine-scoring answer sheets 1½c each; specimen set 20c. Published by the Bureau of Educational Research and Service, State University of Iowa, Iowa City, Iowa.

Iowa Placement Examinations, Foreign Language Aptitude—Form M, constructed by G. D. Stoddard; revised by Grace Cochran, J. R. Nielson, and D. B. Stuit. Revised 1941. For college students. Time, 45 minutes. Hand- and machine-scored. $4.00 per 100; machine-scoring answer sheets 2c; specimen set 20c. Published by the Bureau of Educational Research and Service, State University of Iowa, Iowa City, Iowa.

Iowa Placement Examinations, Physics Aptitude—Form M, constructed by G. D. Stoddard and C. J. Lapp; revised by C. J. Lapp and D. B. Stuit. Revised 1941. For college students. Time, 50 minutes. Hand- and machine-scored. $4.00 per 100; machine-scoring answer sheets 1½c each; specimen set 20c. Published by the Bureau of Educational Research and Service, State University of Iowa, Iowa City, Iowa.

Kansas Spelling Test, by H. E. Schrammel, O. M. Rasmussen, and Wayne Gordon. 1941. Test I for grades 1 to 3; Test II for grades 4 to 6; Test III for grades 7 to 9. Time, 15 minutes. Forms A and B, 50c per 25; specimen set 15c. Published by the Bureau of Educational Measurements, Kansas State Teachers College, Emporia, Kansas.

Language Essentials Tests, by Vera Davis and H. E. Schrammel. 1941. For grades 4 to 8. Time, 30 minutes. Forms A and B, $1.00 per 25; $4.00 per 100; $36.00 per 1000; specimen set 25c. Published by the Educational Test Bureau, Minneapolis, Minnesota.

McDougal General Science Test, by Clyde R. McDougal. 1941. For high school students. Time, 40 minutes each for Test I and Test II. 50c per 25; 2½c each; specimen set 15c. Published by the Bureau of Educational Measurements, Kansas State Teachers College, Emporia, Kansas.

Primary Business Interests, by Alfred J. Cardall. 1941. For high school, college and adult levels. Time, about 20 minutes. Hand- or machine-scored. 5c each; scoring keys 25c; specimen set 35c. Published by Science Research Associates, 1700 Prairie Avenue, Chicago, Illinois.

Primary Reading Tests, by Albert G. Reilley. 1941. For grade 1. Form B, 85c per 25. Published by Houghton Mifflin Company, 2 Park Street, Boston, Massachusetts.

Recreation Inquiry, by Richard Wilkinson and Sidney L. Pressey. For high school and college students. Time, about 50 minutes. $1.00 per 25; $3.00 per 100; specimen set 15c. Published by the Psychological Corporation, 522 Fifth Avenue, New York City.

Wiksell-Filkin Library Instructional Tests, by Wesley Wiksell and Mary Filkin. 1941. For high school and college students. 25 separate tests each on a different subject. $3.75 per 25 complete batteries. Published by the Acorn Publishing Company, Rockville Centre, Long Island, New York.

Wilson Scales of Stability and of Instability, by Matthew H. Wilson. 1941. For junior and senior high school and college students, and adults. Time, 20 to 30 minutes. $1.15 per 25; 5c each; specimen set 15c. Published by the Bureau of Educational Measurements, Kansas State Teachers College, Emporia, Kansas.

MEASUREMENT ABSTRACTS*

Baxter, Brent. "An Experimental Analysis of the Contributions of Speed and Level in an Intelligence Test." *Journal of Educational Psychology,* XXXII (1941), 285-96.

Three measurements were taken of performance on an intelligence test: speed, the time to complete the entire test; power, the number of items correct at the end of a given time; and level, the number of items correct with unlimited time. Speed and level are uncorrelated. "Speed and level contribute the entire variance of power." Level is more important than speed in determining college grades. When the students are tested individually, prediction (of Army Alpha scores, of college aptitude test scores, and of college grades) "through the combination of speed and level in multiple correlation is greater than that possible" with the more usual scores of power. When the students are tested in groups, this superiority vanishes. *H. M. Wolfle.*

Carroll, John B. "A Factor Analysis of Verbal Abilities." *Psychometrika,* VI (1941), 279-308.

A multiple-factor analysis was made of a battery of 42 tests of verbal abilities administered to 119 college adults. Where necessary, the distributions of test scores were normalized before the inter-test correlations were computed. Thurstone's M (Memory or Rote Learning) factor has been confirmed, but his V (Verbal Relations) factor seems to have been split into two or possibly three factors, C, J, and G. His W (Word Fluency) factor has been split into two factors, A and E. The C factor seems to represent the richness of the individual's stock of linguistic responses, and the J factor seems to involve the ability to handle semantic relationships. No satisfactory interpretation can as yet be made of the G factor. The A factor seems to correspond to the speed of association for common words where there is a high degree of restriction as to appropriate responses. The E factor is described as an associational facility with verbal material where

*Edited by Professor Forrest A. Kingsbury.

the only restriction is that the responses must be syntactically coherent. The new factors are: F, facility and fluency in oral speech; H, facility in attaching appropriate names or symbols to stimuli; and D, speed of articulatory movements. (Courtesy *Psychometrika.*)

Chapanis, Alphonse. "Notes on the Rapid Calculation of Item Validities." *Journal of Educational Psychology,* XXXII (1941), 297-304.

Several shortcuts in the estimation of item validities by means of the biserial correlation coefficient are suggested. When one is not interested in inter-test comparisons, the constant $\frac{i}{\sigma_T}$ may be eliminated. By employing the same class interval and assumed mean i may be omitted and the means in the formula replaced by deviations of guessed means. Moreover, if the range of abilities tested is homogeneous, z is unnecessary. Formulae facilitating transformation of reduced coefficients are presented in case inter-test comparisons are later desired. *V. Brown.*

Crissy, William J. E. "A Reply to an Examinee's Reactions to the National Teacher Examinations." *Journal of Higher Education,* XII (1941), 484-487.

This article takes each of the particular criticisms in turn and indicates how each problem of test construction was handled in compiling the National Teacher Examinations. The advantages of the particular test form used are listed and year-to-year comparability of test scores is claimed. The construction of test items and scoring key is briefly outlined. The purpose of the examinations and the use of the test results are considered and precautions taken in this area are pointed out. *D. A. Peterson.*

Ferguson, George A. "The Factorial Interpretation of Test Difficulty." *Psychometrika,* VI (1941), 323-30.

This paper discusses the influence of test difficulty on the correlation between test items and between tests. The greater the difference in difficulty between two test items or between two tests, the smaller the maximum correlation between them. In general, the greater the number of degrees of difficulty among the items in a test or among the tests in a battery, the

higher the rank of the matrix of intercorrelations; that is, differences in difficulty are represented in the factorial configuration as additional factors. The author suggests that if all tests included in a battery are roughly homogenous with respect to difficulty, existing hierarchies will be more clearly defined and psychological interpretation will be more meaningful. (Courtesy *Psychometrika.*)

Flanagan, John C. "A Preliminary Study of the Validity of the 1940 Edition of the National Teacher Examinations." *School and Society,* LIV (1941), 59-64.

Because conclusive validation is not yet possible, this report aims to do no more than to review evidence now available respecting the 1940 National Teacher Examinations. Two meanings of "validity" are distinguished: (1) Do the tests satisfactorily get at the content and mental processes indicated in the outline and specifications by which they were planned? (2) Do they aid in distinguishing between better and poorer teachers as measured by any of numerous criteria of teacher excellence? The former is partially answered by the agreement of the 10 or 12 cooperating experts who critically examined the tests and their specifications, and by intercorrelations between tests of the battery; the latter, by citation of several lines of evidence. One line is student ratings on 49 teachers in 22 systems (at least 2 in each system) who had taken these tests and were chosen so as to reveal considerable spread of scores on the "common examinations." Correlation between ratings and test-scores was .51. Supervisors' ratings on these teachers on a number of items are also cited. The five highest of these correlated around .50 with test-scores. In general, the study shows that the examinations have some predictive value as to the teacher's general effectiveness and desirability, and also points out other significant items. *F. A. Kingsbury.*

Froehlich, Gustav J. "A Simple Index of Test Reliability." *Journal of Educational Psychology,* XXXII (1941), 381-85.

A simple adaptation of the Kuder-Richardson index of test reliability is described, namely:

$$r = \frac{\sigma^2 n - M\,(n - M)}{\sigma^2\,(n - 1)}$$

Since this formula involves only the number of items in the test, the mean of the test scores, and their standard deviation, it is offered as an index of test reliability easily applied by teachers and others who are limited with respect to time and statistical background. An empirical check, using the Wisconsin Achievement Test on some 2000 individuals, shows that reliability coefficients on the total battery and five parts, as computed by this formula, run slightly lower (.017 to .058) than Spearman-Brown *r's*, the two rank orders of the six *r's* being identical. *F. A. Kingsbury.*

Gritten, Frances and Johnson, Donald M. "Individual Differences in Judging Multiple-Choice Questions." *Journal of Educational Psychology,* XXXII (1941), 423-430.

Form A of the *Nelson-Denny Vocabulary Test* was given with instructions not to guess; form B with instructions to answer all questions and to rate each judgment on a confidence scale. Four different achievement scores from Form A were correlated with the confidence and achievement scores from Form B. The results indicated that with instructions not to guess, the more confident subjects will attempt and correctly answer more items, and that the conventional formula, $R-\frac{W}{n-1}$, could properly be called a correction for individual differences in confidence. *V. Brown.*

Haggerty, Lida Harmer. "An Empirical Evaluation of the Accomplishment Quotient. A Four Year Study at the Junior High School Level." *Journal of Experimental Education,* X (1941), 78-90.

"The AQ is a distinctly unreliable measure." This conclusion was reached after studying data for 163 subjects over a four-year period. Intelligence was measured eight times, using four different tests, and achievement four times, combining Forms V and W of the New Stanford Achievement Test. "There is a mean inter-*r* of only .35 for the eight AQ distributions," in spite of duplication among the measures. "Large numbers of pupils (seem to) achieve up to capacity or fall below capacity without the slightest change in their actual work by merely changing from one accepted test to another in measuring intelligence."

A composite of all achievement scores correlates .94 with a composite of all intelligence scores. *H. M. Wolfle.*

Hartmann, George W. "A Critique of the Common Method of Estimating Vocabulary Size, Together with Some Data on the Absolute Word Knowledge of Educated Adults." *Journal of Educational Psychology,* XXXII (1941), 351-358.

Vocabulary estimates based upon samples of varying size drawn from the same dictionary were found to be fairly stable, although results indicated that less than fifty words do not yield an accurate measure. When samples chosen from dictionaries of varying size were compared, vocabulary estimates were discovered to be dependent upon the size of the dictionary. Commonly accepted estimates need upward revision, for the present study demonstrated that the recognition vocabulary of the average undergraduate is in excess of 200,000 words. *V. Brown.*

Horst, Paul, and collaborators. *The Prediction of Personal Adjustment.* New York: Social Science Research Council, pp. xii+455. 1941.

This monograph is a study of the logic and methodology of the prediction of personal adjustment, prepared under the supervision of the Committee on Social Adjustment of the Social Science Research Council. It is oriented primarily with studies in prediction of adjustments in four fields, namely, school success, vocation, marriage, and crime, with a supplementary memorandum on problems of prediction in the national defense program. Five supplementary studies by collaborators are included dealing with case-study techniques, mathematical and tabulation techniques, reduction of number of variables (factor grouping), combining and weighting measures, and five mathematical problems. In the systematic section, detailed descriptions of tests and other instruments are omitted, and the major methodological aspects of the prediction problems are summarized and analyzed in order. A chapter is devoted to suggestions for research projects in the prediction of individual behavior. *F. A. Kingsbury.*

Jackson, R. W. B. "Some Difficulties in the Application of the Analysis of Covariance Method to Educational Problems." *Journal of Educational Psychology,* XXXII (1941), 414-422.

Since the analysis of covariance is a statistical method developed mainly for use in another field, it seems inadvisable to apply it without question to educational problems. Although the method has been found very useful, it may be necessary to modify it slightly. Examples are given to illustrate some of the difficulties likely to be encountered in the adoption of this method and to demonstrate their possible solutions. *V. Brown.*

Langsam, Rosalind Streep. "A Factorial Analysis of Reading Ability." *Journal of Experimental Education,* X (1941), 57-63.

The factors involved in reading ability were determined using Thurstone's centroid method with rotation of axes. Twenty different subtests from reading and intelligence tests and one from the *Primary Mental Abilities* battery were analyzed. Three of the four factors found in reading tests were identified as being similar in character to three of Thurstone's primary abilities: a verbal factor V, "an ability to deal with verbal material"; a perceptual factor P, which in this material shows up as speed in "perceiving and selecting the correct word from other words offered as possible answers"; and a word factor W, "a fluency in dealing with words." A more tentative factor was "that of seeing relationships." *H. M. Wolfle.*

Lennon, Roger T. "Note on Line of Relation Method of Establishing Age or Grade Norms." *Journal of Educational Psychology,* XXXII (1941), 389-90.

Two methods of establishing age or grade norms are: (1) to find mean scores for successive age or grade groups and pass a norm line through them; (2) to determine empirically the correspondence of scores on the new test with those on a test whose "line of relation" has already been established and interpolate norms on such a line of relation. The author points out the condition under which the two methods yield identical results; namely, when the correlation of scores on each of the

two tests with age is the same. The correspondence method is applicable only when this condition is known to be satisfied. *F. A. Kingsbury.*

McCormick, Thomas Carson. *Elementary Social Statistics.* New York: McGraw-Hill Book Company. pp. 353. 1941.

This elementary statistics textbook is designed primarily for students and workers in sociology rather than in psychology and education. The first part of the book deals with the nature and control of statistical inquiry. The second part is devoted to common statistical procedures: tabulation of distributions, graphs, measures of deviation, correlation techniques, sampling and sampling errors, the significance of differences, and analysis of time series. Statistical tables are also included at the end of the book. *Jane Gilbert.*

McNamara, John Joseph. "A New Method for Testing Advertising Effectiveness Through Eye Movement Photography." *Psychological Record,* IV (1941), 399-460.

In order to test advertising effectiveness one group of readers was asked to leaf through a magazine containing advertising matter. Their eye-movements were photographed with the Purdue Eye-Camera and the mean time spent on each part of an advertisement was recorded. The magazine was advance copy so the readers had no opportunity to see the copy prior to the experiment. Another group was given advertisements which had been cut up into parts and pasted on cardboard in heterogenous order. The length of time the reader required to identify the parts with the whole advertisement was recorded. The probability that the reader would look at the advertisement long enough to identify the advertiser was also computed. The reliability of these techniques was high. The correlation between mean time and probability scores was .48 for combined groups. The effect of magazine position, position on the spread, and cartoons on advertising effectiveness was also determined. *Jane Gilbert.*

Peters, Charles C. and Van Voorhis, Walter R. *Statistical Procedures and Their Mathematical Bases.* New York: McGraw-Hill Book Company. pp. 516. 1941.

This textbook is designed to explain the mathematical origins of statistical formulae in terms that can be understood

by statistical workers having little mathematical background. Many of the Fisher techniques are discussed in relation to theoretical statistics, although the limitations of these techniques as applied in the psychological and social sciences are also indicated. The authors present a brief discussion of calculus and elementary statistical procedures, measures of central tendency, variability, reliability, probability, multiple-factor analysis, curve fitting, partial and multiple correlation, the nature of Chi squared, and the techniques used in controlled experimentation. *Jane Gilbert.*

Remmers, H. H. and House, J. Milton. "Reliability of Multiple-choice Measuring Instruments as a Function of the Spearman-Brown Prophecy Formula, IV." *Journal of Educational Psychology,* XXXII (1941), 372-76.

The hypothesis tested is that the relation between changes in reliability of multiple-choice test-items and changes in number of response alternatives per test-item is predictable by the Spearman-Brown formula. A 60-item, five-alternative, multiple-choice arithmetic test was given to 771 junior high school pupils. Three derivative forms were constructed from this, having respectively four, three, and two alternative answers per item, the eliminated answers having been selected by lot. Four groups, equated as to I.Q., took separate forms of the test. Reliability coefficients of half-test (odd-even) and whole-test both showed regular decrease for the four forms with decreasing number of alternatives, thus supporting the hypothesis within this range of two to five alternative responses. *F. A. Kingsbury.*

Remmers, H. H. and Sageser, H. W. "Reliability of Multiple-choice Measuring Instruments as a Function of the Spearman-Brown Formula, V." *Journal of Educational Psychology,* XXXII (1941), 445-51.

The hypothesis tested is the same as that described in the Remmers and House article, abstracted above. Two equivalent attitude-scales (Remmers & Bues) of 37 items were combined and used in testing attitudes of agreement or disagreement on each of two college practices. Four derivative sets were prepared, providing respectively two, three, five, and

seven degrees of choice. From 87 to 112 university students filled out each of these forms. Each of the two tests was scored twice, once with equal values for all statements, once with items weighted by scale-value. Dividing each of the four sets of scores into the two original 37-item scales, four correlations were found for each set of papers. Corrected for skewness by being transformed into "z" functions, the obtained reliabilities with unweighted scores did not support the hypothesis; but when weighted in terms of the experimentally determined scale values of the scale-items, the data supported the hypothesis. *F. A. Kingsbury.*

Ruch, Floyd L. "The Problem of Measuring Morale." *Journal of Educational Sociology,* IV (1941), 221-228.

At the present time one of the more effective tools developed to give an orderly description of public response is the opinion poll. The two basic problems in public opinion polling are: to get a representative sample, and to get the desired information from every case in the sample. The problem of sampling is discussed and references are given for specific techniques in the field. In the second problem basic types of defects are listed which lower the dependability and accuracy of questions. In conclusion possible additions to the public opinion poll technique in the measurement of morale are discussed. *D. A. Peterson.*

Satterthwaite, Franklin E. "Synthesis of Variance." *Psychometrika,* VI (1941), 309-16.

The distribution of a linear combination of two statistics distributed as is Chi-square is studied. The degree of approximation involved in assuming a Chi-square distribution is illustrated for several representative cases. It is concluded that the approximation is sufficiently accurate to use in many practical applications. Illustrations are given of its use in extending the Chi-square, the Student "t" and the Fisher "z" tests to a wider range of problems. (Courtesy *Psychometrika.*)

Spache, George. "Deriving Comprehension, Rate and Accuracy of Reading Norms for a Short form of the

Metropolitan Achievement Reading Test." *Journal of Educational Psychology,* XXXII (1941), 359-64.

The *Metropolitan Achievement Tests,* although widely used, require a long testing time (about 2¾ hours for Intermediate Partial). The uses to which the test is put indicate that abbreviation of certain tests is preferable to omission of sub-tests, if the time has to be shortened. The method used in abbreviating the Reading Test is described, and correlation coefficients between total grade-score and shortened test raw-scores are cited. Grade-score norms for the short form are given, derived from the regression equations, and also percentile norms for reading accuracy, the latter based on private schools and of uncertain validity for public schools. Validity coefficients for the shortened form are found to be almost as high as the reliability coefficients of the long form. *F. A. Kingsbury.*

Thornton, G. R. "The Use of Tests of Persistence in the Prediction of Scholastic Achievement." *Journal of Educational Psychology,* XXXII (1941), 266-274.

A factor analysis of persistence tests revealed two unrelated factors; one appeared in the shock and pressure tests, the other in the word building and perceptual ability tests. The hypothesis that personality tests have value for predicting scholastic success in proportion to the degree of similarity between tests and classroom situations is suggested to explain the lower correlation between achievement and persistence found in this investigation. A formula utilizing scholastic efficiency and aspiration displayed in a previous school is presented for prediction of grades in a new school. *V. Brown.*

Travers, L. B. "Improving Practical Tests." *Personnel Journal,* XX (1941), 129-33.

This article discusses the advantages of evaluating the personal characteristics of testees while they are undergoing a practical test, rather than while interviewing them. The ratings under actual working conditions are claimed to be more reliable. Sample charts are given for a carpenter's practical test and the rating sheet used with it. *H. M. Wolfle.*

Traxler, Arthur E. "A Study of the Junior Scholastic Aptitude Test." *Journal of Educational Research,* XXXV (1941), 16-27.

The Junior Scholastic Aptitude Test consists of three parts: verbal section, containing five subtests; numerical section, containing three subtests; and an experimental section, not included in pupil's score. A practice booklet is issued several days prior to the examination for the student to work through at his leisure. The administration of the test proper is on a secret basis. Results are reported to the school in terms of a derived score. The reliability, validity, and prognostic value of the test are discussed as indicated by correlations with other tests of academic aptitude, achievement tests, and school marks. "The data are not extensive enough to be conclusive, but it is hoped that they will be of some assistance in appraising and using this new test." *D. A. Peterson.*

Wherry, Robert J. "An Extension of the Doolittle Method to Simple Regression Problems." *Journal of Educational Psychology,* XXXII (1941), 459-464.

This article describes a new method for solving simple regression constants which the author has found very successful in teaching beginners. It is shorter not only because it involves fewer arithmetical operations, but also because it is more systematic. Other advantages claimed are that the checks are more certain and convincing, and once the beginner has mastered the technique involved in simple correlation, he is immediately able to solve multiple correlation constants in precisely the same manner with little further training. *V. Brown.*

Winetrout, Kenneth. "The National Teacher Examinations, 1941." *Journal of Higher Education,* XII (1941), 479-484.

The writer gives a brief, general description of the length of the examinations and the method of giving them. This is followed by criticisms both adverse and appreciative of the construction of the examinations and potential use of test results obtained. It is suggested that if the question form were

varied in the long testing periods (the examinations, total duration being twelve hours, are restricted to the use of the multiple-choice question form), a more adequate examination program might be obtained. The author believes that one section was concerned with measurement of attitudes rather than capacities and would substitute a three-point (conservative, liberal, radical) rating scale for the right-wrong classification used at present. Wording of questions, sectional influence, and factual emphasis are also discussed by the writer, who recently took the examinations. *D. A. Peterson.*

Young, Gale. "A Note on Multidimensional Psychophysical Analysis." *Psychometrika,* VI (1941), 331-33.

On viewing Thurstone's psychophysical scale from the point of view of the mathematical theory of one-parameter continuous groups, it appears that a variety of different psychological or statistical assumptions can all be made to lead to a scale possessing similar properties, though requiring different computational techniques for their determination. The natural extension to multi-dimensional scaling is indicated. (Courtesy *Psychometrika.*)